AD_____

Award Number: W81XWH-04-1-0294

TITLE: DNA Repair and Ethnic Differences in Prostate Cancer Risk

PRINCIPAL INVESTIGATOR: Radoslav Goldman, Ph.D.

CONTRACTING ORGANIZATION: Georgetown University
Washington DC 20057

REPORT DATE: March 2008

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-03-2008	Annual Summary	15 Feb 04 – 14 Feb 08

4. TITLE AND SUBTITLE
DNA Repair and Ethnic Differences in Prostate Cancer Risk

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0294

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Radoslav Goldman, Ph.D.

E-Mail: RG26@GEORGETOWN.EDU

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Georgetown University
Washington DC 20057

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT: Prostate cancer is the most common lethal tumor among US males and is particularly high in African Americans. This study evaluates DNA repair in a study of 240 prostate cancer patients and 240 healthy controls matched on age and race in a 50% African American population. Low DNA repair correlates with increased risk of certain cancers but prostate cancer was not yet examined. We hypothesize that low DNA repair, especially in African American men, contributes to increased risk of having prostate cancer. To evaluate this hypothesis, we quantify DNA repair capacity in blood cells using comet assay and evaluate how this repair capacity is related to genetic variants in OGG1 and XRCC1 DNA repair genes. Genetic variants of OGG1 and XRCC1 with a decreased DNA repair capacity were previously identified. This means that a portion of the general population carrying the 'at risk' variant might be at higher risk of developing prostate cancer.This pilot study is expected to fill important gaps in our understanding of prostate cancer etiology, produce new hypotheses which can be tested in an expanded prostate cancer study, focus prostate cancer prevention in a new direction, and help design better cancer prevention and treatment strategies.

15. SUBJECT TERMS
DNA repair, prostate cancer, comet assay, single nucleotide polymorphism, molecular epidemiolgy

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	124	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

INTRODUCTION:

Despite the fact that prostate cancer is the most common tumor among US males, relatively little is known about the causative mechanisms (1). We known that the incidence increases with age, varies by geography and by ethnicity, and is higher among men whose father or brother had the disease. These factors, however, are not sufficient for identification of men with increased susceptibility. African American males are particularly susceptible with highest rates of prostate cancer world-wide and about twice the rates of Caucasian Americans which holds true for every age group, clinical stage, and histological classification; this is even more striking in view of the lower screening among African Americans (2). It is not known what causes higher rates in African Americans, but some studies suggest differences in cancer biology (3). Recent studies show that rates in Africa are much higher than previously considered and comparable to the rates of African Americans (4;5). And new analyses estimate that the heritable contribution to prostate cancer risk including high and low penetrant genes is as high as 42% (CI 29%-50%) (6) in spite of the acknowledged contribution of the environment based on migrant studies (3). As very little is known about the genetic modifiers of prostate cancer risk, establishing new biomarkers would greatly benefit the field of prostate cancer prevention and surveillance, as well as advance our understanding of ethnic health disparities.

Our hypothesis is that prostate cancer risk and ethnic risk differences are related to interindividual variability in DNA repair. In this study, we compare DNA repair capacity of cancer cases and controls and stratify the results by race (African American and Caucasian men). DNA repair capacity was quantified by two phenotypic assays, comet assasy and mutagen sensitivity. In addition, we performed a comparison of 13 estrogen metabolites in urine of patients enrolled in our study. The hypothesis for this supplemental study was that estrogenic signaling through the estrogen receptor beta receptor controls growth of prostate tissue and protects from cancer.

DNA Repair and Comet Assay: DNA repair consists of two major categories, excision repair (base excision repair and nucleotide excision repair) and recombination repair (homologous and non-homologous) (7). In prostate, mismatch repair genes have lower activity and are down regulated in cancer cell lines (8) and tumor tissue (9). This repair pathway, associated with hereditary nonpolyposis colorectal cancer, could be also associated with prostate cancer (10;11). Numerous polymorphisms in the DNA repair genes have been identified and are likely to contribute to cancer risk (12). But two functional polymorphisms, *OGG1* and *XRCC1*, are particularly relevant to this study. Prostate cancer is related to chronic inflammation (13) and oxidative DNA damage (14); and lycopene, vitaminE, and other antioxidants are suggested protective agents (15). Both *OGG1* and *XRCC1* repair oxidative DNA damage, and both genes have been recently associated with prostate cancer risk in case control studies (16;17). It is therefore plausible that variability in the DNA repair efficiency contributes to prostate cancer susceptibility. To capture the variation in this complex pathway, we propose phenotypic quantification by comet assay.

Comet or single cell gel electrophoresis assay (SCGE) quantifies unwinding of nuclear DNA under alkaline (pH>13) electrophoresis conditions (18). This provides a measure of DNA damage reflecting the presence of alkali labile sites, single and double strand breaks (19). The kinetic of comet disappearance provides a simple and robust measure of DNA repair increasingly popular in human biomonitoring (18). The assay can be used for quantification of DNA damage and repair in a variety of cells including short-term cultured human lymphocytes. This approach was used recently in three pilot studies of breast, cervical, and lung cancer and demonstrated the potential of comet assay to identify cancer-prone individuals in the general population (20). The largest of the studies examined lymphocytes of 160 lung cancer patients and 180 controls by comet assay. High DNA damage (OR 4.2; CI 2.2-7.4) and deficient DNA repair (OR 2.1; CI 1.1-4.0) following exposure to bleomycin were independent predictors of cancer risk (21). Bleomycin is a radio mimetic inducing oxidative DNA damage, a good model for the suspected prostate carcinogenesis. This would be the first study to use comet assay as a DNA repair capacity screen in prostate cancer risk.

Significance: This is one of the few molecular epidemiology studies that examined phenotype of DNA repair capacity (comet assay and mutagen sensitivity) as a biomarker of prostate cancer risk. A number of lines of evidence suggest that variation in DNA repair may be an important determinant of prostate cancer risk (9;14;16;17). Ethnic differences in the DNA repair capacity are evaluated. The proposal is innovative because the proposed biomarker was not examined in prostate cancer. If comet assay correlate with prostate cancer risk, it could serve as readily obtainable biomarkers to identify men with increased risk of prostate cancer and focus prevention and intervention strategies. The phenotypic biomarkers could be used to better characterize genotoxic insults leading to cancer risk (improved risk models). The budget constraints prevent us from investigating a larger population, but the preliminary results from this research will be used to seek funding of an expanded study testing further hypotheses and associations. Elucidating mechanisms of the early stages of prostate carcinogenesis would have an immediate impact for prevention and surveillance. Better prevention strategies (including chemoprevention) could be designed and tested based on the identified targets. And new hypotheses focusing on the genetic and environmental factors associated with prostate cancer risk could be formulated and evaluated.

BODY:

This is a case-control study of prostate cancer risk which collected blood sample, urine, and data on prostate cancer patients and age and race matched controls in order to examine contribution of DNA repair capacity to cancer risk. The goal was to examine DNA repair capacity as a risk factor in prostate cancer development and to compare the DNA repair phenotype in African American and Caucasian populations. To this end, we recruited 131 cases and 231 controls at Georgetown University Hospital (GU) and Veterans Administration Hospital (VA), Washington, DC. Epidemiological data, clinical data, blood sample, and urine were obtained and comet assay phenotype was evaluated in white blood cells exposed to ionizing radiation following an overnight storage of whole blood at $4^{o}C$ (22). These markers were correlated with prostate cancer risk independently and in combination. In addition, we carried out mass spectrometric quantification of 14 metabolites of estrogen in prostate cancer cases and controls to evaluate the influence of estrogenic signaling on prostate cancer risk.

Patient recruitment and data collection: The patient enrollment and data/sample collection was carried out at GU and VA hospitals, Washington, DC. The GU clinics see similar volume of patients as the VA hospital, but the patient population at GU is about 70% Caucasian while the VA prostate patient population is about 70% African American. The patients enrolled in this study are adult residents of the greater Washington, DC area including Maryland and Virginia suburbs. We enrolled all eligible patients that cover the full spectrum of tumor stage and grades. All subjects were briefly informed about the study by the attending physician and referred for further information to a study coordinator. The interviewer briefly described the study, answered patient's questions, and obtained informed consent of interested participants. To be eligible, patients must have been at least 18 years of age and have not previously been diagnosed with any other cancer besides non-melanoma skin cancer at the time of enrollment. All participants were enrolled prior to radiation, surgery, or chemotherapy. We enrolled two groups of controls. One group consisted of men coming for urological examination and confirmed by biopsy to be cancer free at the time of enrollment. Approximately 33% of the men have cancer at biopsy; the remaining patients are confirmed to be cancer free and serve as a control group. This control group consists primarily of men with benign prostatic hypertrophy (BPH) and some other non-malignant urologic conditions including prostatitis. The control group of healthy men without apparent urologic problems consists of visitors accompanying other patients to the hospital. A free PSA test is provided for the controls without a verifiable recent result. We excluded spouses and blood relatives of patients to avoid overmatching on genetic factors. The study coordinator identified potential candidates, investigated their willingness to participate, and screened for eligibility. The study coordinator worked from a table of enrolled cases and frequency-matches the eligible controls. The study coordinator obtained informed consent, questionnaire data (including dietary questionnaire), retrieved clinical information, and assisted with collection of biological specimen. Each subject provides a single 45 cc blood sample drawn into pre-labeled vacutainer glass tubes. Upon receipt into the laboratory, the samples were verified against the shipping papers and logged into our repository database. We collected two red top tubes (no preservative), two green top tubes (sodium heparin), a yellow top tube (ACD), and one purple top tube (EDTA). Urine and saliva were collected according to standard procedures and frozen for future studies as needed. One fresh aliquot of heparinized blood was used immediately for DNA repair assays as described below. The remaining specimens were processed according to a standard protocol and aliquots of bar-coded serum, plasma, and white blood cells were frozen at $-80^{o}C$ for further studies. The questionnaire asked about demographic information, reproductive history, tobacco use, alcohol consumption, general medical history and family history, occupational exposures, residential history, exercise, and education **(see Appendix)**. This information was entered into an Epi Info databases using double entry forms. Additional clinical information including stage and grade of tumors was extracted from medical records. Thus, we created a fully annotated repository of specimen of 362 participants (131 cases and 231 controls). From this total, 55

cases and 124 controls were evaluated by comet assay. The characteristics of this population are described in **Table 1**.

Table 1	Cases		Controls						p values* cases vs all controls	p-values cases vs biopsy controls
			All		Biopsy		Healthy			
N	55		124		71		53			
Age										
mean	65.2		63.4		65.1		61.3		0.2097	0.9538
st dev	7.4		9.1		9.1		8.6			
missing	3		14		11		3			
PSA										
mean	8.0		3.9		5.3		2.1		0.0035	0.1260
median	5.6		3.8		4.6		1.4			
st dev	14.5		2.9		2.7		1.9			
PSA>4	40	72.73%	57	46.34%	48	67.61%	9	17.31%	0.0011	0.5344
PSA≤4	15	27.27%	66	53.66%	23	32.39%	43	82.69%		
missing			1				1			
Race										
white	39	72.2%	88	71.0%	37	52.1%	51	96.2%	0.4624	0.1181
afr.-amer.	15	24.2%	32	25.8%	30	42.3%	2	3.8%		
other	8	17.0%	4	3.2%	4	5.6%				
BPH[1]									0.8483	0.9463
yes	20	37.7%	42	36.2%	26	37.1%	16	34.8%		
no	33	62.3%	74	63.8%	44	62.9%	30	65.2%		
missing	2		8		1		7			
NSAID									0.8411*	0.9784
occasionally	21	47.7%	51	45.9%	28	47.5%	23	44.2%		
daily	23	52.3%	60	54.1%	31	52.5%	29	55.8%		
missing	11		13		12		1			
Smoking									0.2016*	0.9731
never	21	46.7%	37	31.9%	28	44.4%	9	17.0%		
ex	20	44.4%	63	54.3%	29	46.0%	34	64.2%		
current	4	8.9%	16	13.8%	6	9.5%	10	18.9%		
missing	10		8		8					

* Chisq for categorical variables and ttest for continuous variables. [1] BPH information from the medical records. Those without a medical record have the BPH status evaluated from the Prostate health section of the questionnaire.

Comet Assay: Comet assay can be used to quantify DNA damage and repair in a variety of cells including short-term cultured human lymphocytes and prostate cancer cells. Comet or single cell gel electrophoresis assay (SCGE) quantifies unwinding of damaged nuclear DNA under alkaline (pH>13) electrophoresis conditions. This provides a measure of DNA damage reflecting the presence of alkali labile sites, single and double strand breaks. The kinetic of comet disappearance provides a measure of DNA repair increasingly popular in human biomonitoring. Our method builds on the protocol of Singh, et. al. (23) as described by Schmezer et al. (24). We tested a number of experimental conditions comparing the following conditions: 1. Exposure of cells in suspension or cells embedded in agarose; 2. Exposure of short term cultured isolated lymphocytes and exposure of whole blood stored overnight at 4^{o}C; 3. Exposure to bleomycin (a radiomimetic) and ionizing radiation (0-10Gy); and 4. Quantification of repair kinetic at various time points between 0 and 45 minutes.
Results of the optimization are described in **Appendix 2** and briefly summarized below.

A typical result of an exposure of white blood cells to ionizing radiation is shown in **Figure 1**.

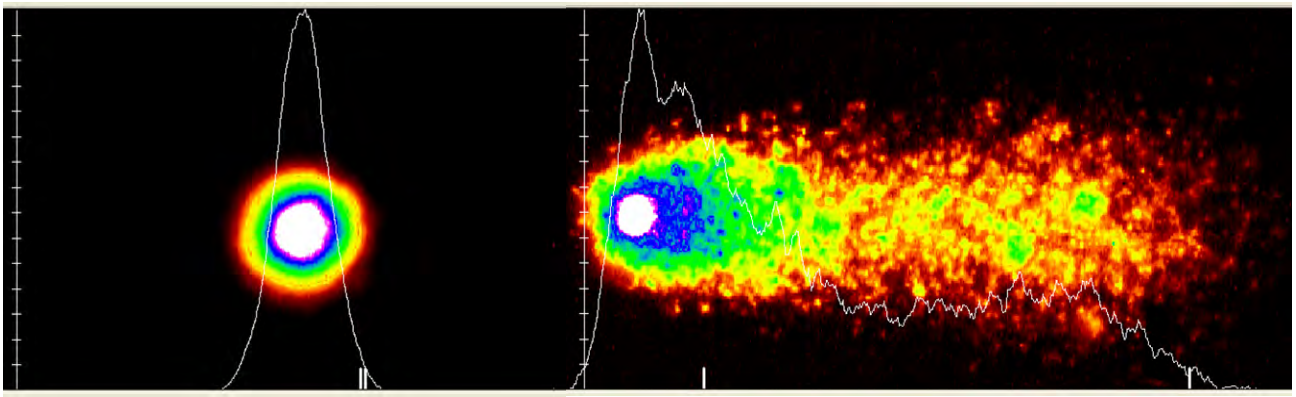


Figure 1. **Control cell** **Cell exposed to 9 Gy Ionizing Radiation**

The example shows images of two cells from an experiment exposing whole blood embedded in agarose to 9 Gy of ionizing radiation; this dose was selected as the ideal exposure to assess DNA repair kinetic white blood cells. Nuclei of control cells (prior to exposure) migrate in the electric field as a compact sphere and show minimal percentage of DNA in the tail region. Nuclei of exposed cells unwind in the electric field and form a tail which can be visualized by the ethidium bromide staining and quantified. Only a small portion of the DNA in the damaged nucleus remains in the head region (the circle at the left side of the image). The intensity of staining is color coded with highest intensity in white and lowest intensity in red. The kinetic of repair (disappearance of the tail induced by ionizing radiation) can be followed and leads to an almost complete repair of the damage within 45 minutes. The above comparison was carried out on cells embedded in agarose because our results show that DNA repair kinetic following exposure to bleomycin or ionizing radiation does not differ between cells exposed in solution or embedded in agarose on a microscopic slides **(Figure 2)**.

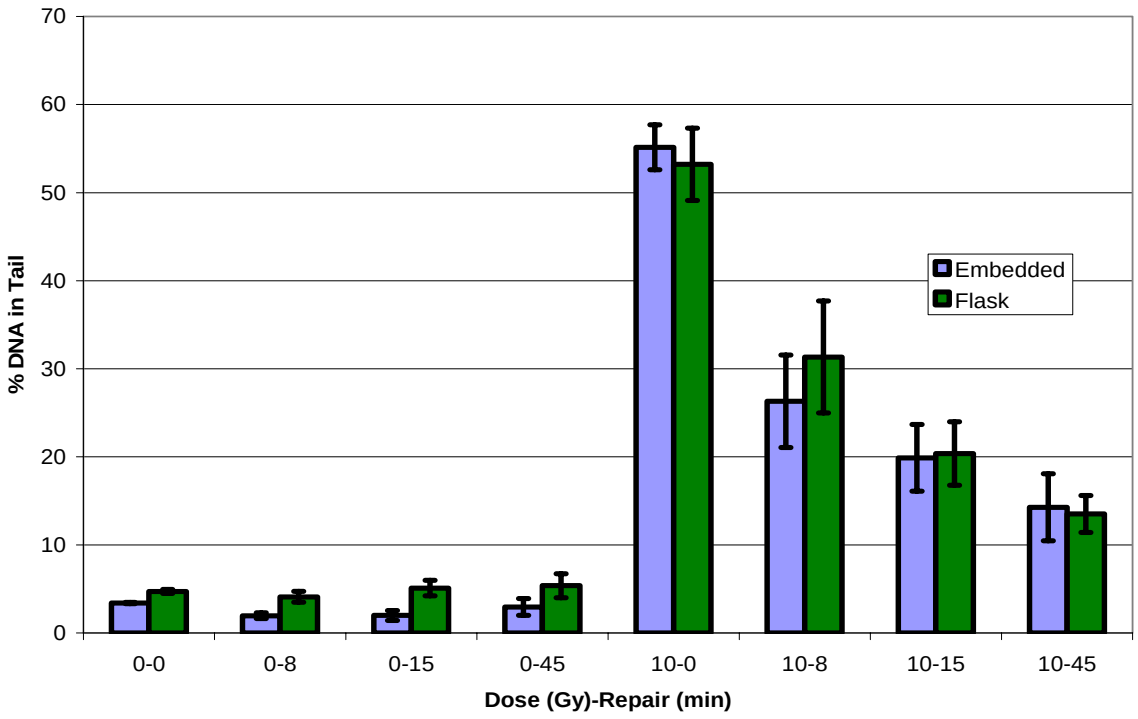


Figure 2. Comparison of cultured lymphocytes exposed to IR (10 Gy) in culture media (green bar) and embedded in agarose (blue bar).

The embedding of cells prior to exposure facilitates the measurement of the DNA repair kinetic; the timing of the repair is more accurate when the embedding step (which requires addition of cells in warm agarose to the microscopic slide) is carried out prior to exposure (see protocol below). The analysis of whole blood simplifies the procedure. Red blood cells do not have nuclei and are not analyzed by this procedure. The white blood cells are used as a surrogate for estimation of DNA repair capacity in prostate and separation of lymphocyte subpopulation is not necessary.

We were also interested in testing of cryopreserved lymphocytes which would allow us to avoid testing of patient samples at inconvenient times. Experiments with cryopreserved cells (slow freezing in 90% FBS with 10% DMSO) showed a significantly higher background DNA damage and slower kinetic of repair compared to fresh cells (data not shown). Based on these experiments, we selected to work with fresh blood. The storage of blood at 4^{o}C was chosen to standardize the procedure and to allow the experiments to start in the morning and be carried out to completion in one day. We typically complete the experiments one day, store dried slides and stain and evaluate rehydrated slides at a later convenient time (typically second day).

The dose of 9 Gy was selected based on dose response experiments which showed an appropriate DNA damage (approximately 50%) immediately following exposure to 9 Gy. The repair kinetic was measured at 15 minutes and at 45 minutes after exposure because the repair seems to be biphasic. The faster kinetic (presumably single strand break repair) is assessed at 15 minutes; the slower kinetic (presumably double strand break repair) is assessed at 45 minutes.

The experimental protocol used for exposure of patient samples is presented below.

1) Coat microscopic slide with 0.75% normal melting point agarose (NMPA), solidify on ice for 5 min
2) Add cell suspension to 0.7 % low melting point agarose (LMPA) at 37°C and form a layer of cells suspended in LMPA (75 µl) on top of the NMPA coated slide
3) Expose embedded cells to 9 Gy ionizing radiation with slides kept at 4^{o}C
4) Allow cells to repair DNA damage for 15 and 45 minutes in RPMI media at 37^{o}C
5) Dip the preparation in cold alkaline (pH 10) lysing solution (4°C) for 3 hours (10 mM Tris, 100 mM EDTA, 2.5 mM NaCl, 1% sodium sarcosinate, 1% Triton X-100, 10% dimethylsulfoxide)
6) Transfer the preparations from lysing solution to alkaline electrophoresis buffer (1 mM EDTA, 300 mM NaOH, pH13) for 40 minutes to unwind DNA
7) Separate DNA in a horizontal gel electrophoresis unit filled with the same buffer for 25 minutes at 4°C by alkaline electrophoresis using 0.92 V/cm and 300 mA current
8) Neutralized slides in 400mM Tris, pH 7.5, fix with methanol, and wash with distilled water
9) Stain with 0.01% ethidium bromide
10) Acquire 100 images per dose/time point (50 cell images per slide, 2 slides) using a fluorescent microscope with a CDD camera (Olympus) and evaluate average fluorescent intensity in the head (intact nuclear DNA) and tail (damaged DNA) using comet imaging software (Loats Associates, Westminster, MD). This imaging system was purchased by Lombardi Comprehensive Cancer Center and installed in our laboratory. The parameter "Percent DNA in Tail" was used for all calculations. The means and standard deviations for each dose (Gy)- repair (min) point were calculated from these 100 measurements.

Comet assay results in the prostate case-control study: The optimized conditions were used to examine the prostate cancer patients and controls. We completed comet assay on 179 study participants (25% African American). The results of a t-test comparison show that there is no difference in the DNA repair between cases and controls, neither all controls nor the subgroups of biopsy and/or healthy controls **(Table 2)**.

Table 2	Mean	Median	ttest
Cases N=55			
Tail 0	38.9	40.9	
Tail 15	24.1	24.3	
Tail 45	14.0	13.9	
Repair 0-15	37%	39%	
Repair 15-45	27%	26%	
Repair 0-45	64%	65%	
Biopsy controls N=71			
Tail 0	38.3	39.0	0.8010
Tail 15	23.0	22.6	0.4500
Tail 45	14.4	13.8	0.6976
Repair 0-15	40%	41%	0.3182
Repair 15-45	23%	23%	0.0834
Repair 0-45	62%	64%	0.5247
Healthy controls N=53			
Tail 0	39.3	42.0	0.8658
Tail 15	22.4	23.1	0.3378
Tail 45	14.2	14.5	0.8599
Repair 0-15	42%	42%	0.0955
Repair 15-45	22%	23%	0.1238
Repair 0-45	64%	65%	0.8554

A multivariate regression analysis adjusted for age and PSA shows also that there is no difference in DNA repair by case control status as determined by the comet assay.

Table 3 Adjusted for psa, age	All controls				Adjusted for psa, age	Biopsy controls			
	OR	95% CI		p value		OR	95% CI		p value
Tail 0	1.0	1.0	1.0	0.8419	Tail 0	1.0	1.0	1.1	0.2880
Tail 15	1.0	1.0	1.1	0.4836	Tail 15	1.0	1.0	1.1	0.3019
Tail 45	1.0	0.9	1.0	0.5069	Tail 45	1.0	0.9	1.1	0.7816
Repair 0-15	0.4	0.0	3.7	0.4165	Repair 0-15	0.6	0.0	8.9	0.6866
Repair 15-45	10.0	0.7	150.5	0.0968	Repair 15-45	11.0	0.4	286.6	0.1484
Repair 0-45	3.0	0.2	39.7	0.3959	Repair 0-45	4.7	0.2	100.0	0.3186

A similar result was obtained when the data was stratified by race; neither the African American nor Caucasian subgroups show case-control differences in DNA repair capacity. However, whm we examined the association of the comet DNA repair phenotype with race, we observed an association between a slower DNA repair phenotype and African-American race **(Table 4)**.

Table 4. Slower repair in African American controls

	All Controls					Biopsy controls only				
	white (N=88)		AA (N=32)		t test	white (N=37)		AA (N=30)		t test
	mean	SD	mean	SD	p value	mean	SD	mean	SD	p value
Tail 0	38.79	12.49	38.99	10.00	0.98	38.76	11.86	38.12	9.55	0.74
Tail 15	22.38	8.22	24.05	7.39	0.43	22.60	7.53	23.83	7.36	0.64
Tail 45	13.85	6.96	15.83	5.49	0.23	13.36	6.06	15.96	5.34	0.11
Repair 0-15	42%	15%	37%	15%	0.12	41%	12%	37%	15%	0.22
Repair15-45	23%	16%	20%	15%	0.24	25%	12%	20%	16%	0.17
Repair 0-45	65%	15%	58%	17%	**0.03**	66%	11%	57%	16%	**0.01**

This DNA repair difference was clearly observed in the biopsy control group; we did not have a sufficient number of African Americans in the healthy control group. The association was no longer observed in the group of prostate cancer cases **(Table 5)**.

Table 5. Differences in comet results by race in cancer cases

	Cases				
	white (N=39)		AA (N=15)		t test
	mean	SD	mean	SD	p value
Tail 0	39.62	14.92	36.90	9.23	0.53
Tail 15	24.27	10.34	23.63	6.84	0.83
Tail 45	14.10	7.05	13.74	4.72	0.79
Repair 0-15	37%	18%	36%	8%	0.86
Repair 15-45	27%	14%	26%	8%	0.96
Repair 0-45	64%	12%	63%	10%	0.77

We examined also the association of the use of NASAIDs, body mass index, benign prostatic hypertrophy, diabetes, alcohol use, and age on the comet DNA repair phenotype and did not find any positive results (data not shown). The above results were summarized in an article that is currently submitted for review.

Mutagen sensitivity phenotype: In addition to the comet assay, we evaluated a small pilot of cases (n=15) and healthy controls (n=50) by mutagen sensitivity assays **(Table 6)**.

Table 6	Cases		Controls		
	N=17	%	N=50	%	χ^2
Mean breaks per cell	48.76		35.62		
Median breaks per cell	48		35		
Standard deviation	14.36		17.87		0.0079*
Age (mean)	63.69	5.29	61.23	10.46	0.3735*
White	15	88.2%	48	96.0%	0.2431
Other	2	11.8%	2	4.0%	
Age less than 62[1]	8	47.1%	28	56.0%	0.523
Age more than 62	9	52.9%	22	44.0%	

The mutagen sensitivity assay phenotype was evaluated in short term cultured cells exposed to bleomycin as described previously (25). It is assumed that this phenotype also reflects DNA repair but there is a number of other cellular processes involved in the strand break phenotype. It is not clear at this poiunt whether the difference in the mutagen sensitivity and comet assay phenotypes observed in our study is due to the differences in exposure (bleomycin vs ionizing radiation) or due to a different mechanism underlying the observed phenotype.

Metabolites of estrogen as risk factors for prostate cancer: In addition to DNA repair phenotypes, we began to utilize the established repository for additional assay evaluating the influence of estrogenic signaling on the prostate cancer risk. This study is still ungoing and the results are very preliminary. We used mass spectrometric assays described previously (26;27) to quantify 14 metabolites of estrogen in the urine of cases and controls **(Table 7)**. At the first pass, we did not observe any case-control differences. The results are currenly adjusted by urinary creatinine and analyzed in detail. The results will be summarized for in a publication as we complete the analyses.

Cases	16KE2	E3	16aE1	16epiE3	17epiE3	3ME1	2ME1
mean	222.46	1784.05	391.88	62.17	36.64	72.01	43.80
sd	460.79	7399.22	1761.56	107.41	107.47	159.09	63.09
median	81.76	192.27	42.27	34.60	9.48	16.66	25.84
Controls	16KE2	E3	16aE1	16epiE3	17epiE3	3ME1	2ME1
mean	345.15	1800.79	614.99	82.37	35.29	72.66	52.50
sd	1588.91	7479.75	4946.80	357.44	160.06	247.61	144.11
median	114.82	295.93	58.96	38.60	11.34	22.38	31.74
ttest	0.51	0.99	0.70	0.63	0.95	0.98	0.62
Cases	4ME1	2ME2	E1	4ME2	E2	2OHE1	2OHE2
mean	22.40	17.05	239.62	16.52	197.65	116.43	28.33
sd	63.76	55.70	663.72	56.35	816.08	139.55	39.72
median	7.28	6.14	89.48	4.72	61.75	73.67	19.80
Controls	4ME1	2ME2	E1	4ME2	E2	2OHE1	2OHE2
mean	20.92	13.76	169.01	9.45	94.68	129.16	32.57
sd	58.15	29.96	181.58	16.49	120.71	131.81	44.00
median	7.81	6.52	112.14	5.97	50.77	76.42	17.59
ttest	0.87	0.60	0.28	0.21	0.19	0.53	0.50

Reference List

1. Kolonel, L. N., Nomura, A. M., and Cooney, R. V. Dietary Fat and Prostate Cancer: Current Status. J Natl Cancer Inst. 3-3-1999;91(5):414-28.

2. Hoffman, R. M., Gilliland, F. D., Eley, J. W., Harlan, L. C., Stephenson, R. A., Stanford, J.

L., Albertson, P. C., Hamilton, A. S., Hunt, W. C., and Potosky, A. L. Racial and Ethnic Differences in Advanced-Stage Prostate Cancer: the Prostate Cancer Outcomes Study. J.Natl.Cancer Inst. 3-7-2001;93(5):388-95.

3. Hsing, A. W. and Devesa, S. S. Trends and Patterns of Prostate Cancer: What Do They Suggest? Epidemiol.Rev. 2001;23(1):3-13.

4. Ogunbiyi, J. O. and Shittu, O. B. Increased Incidence of Prostate Cancer in Nigerians. J.Natl.Med.Assoc. 1999;91(3):159-64.

5. Angwafo, F. F., Zaher, A., Befidi-Mengue, R., Wonkam, A., Takougang, I., Powell, I., and Murphy, G. High-Grade Intra-Epithelial Neoplasia and Prostate Cancer in Dibombari, Cameroon. Prostate Cancer Prostatic.Dis. 2003;6(1):34-8.

6. Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., and Hemminki, K. Environmental and Heritable Factors in the Causation of Cancer--Analyses of Cohorts of Twins From Sweden, Denmark, and Finland. N.Engl.J.Med. 7-13-2000;343(2):78-85.

7. Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. Human DNA Repair Genes. Science 2-16-2001;291(5507):1284-9.

8. Yeh, C. C., Lee, C., and Dahiya, R. DNA Mismatch Repair Enzyme Activity and Gene Expression in Prostate Cancer. Biochem.Biophys.Res.Commun. 7-13-2001;285(2):409-13.

9. Chen, Y., Wang, J., Fraig, M. M., Metcalf, J., Turner, W. R., Bissada, N. K., Watson, D. K., and Schweinfest, C. W. Defects of DNA Mismatch Repair in Human Prostate Cancer. Cancer Res. 5-15-2001;61(10):4112-21.

10. Kolodner, R. D. Mismatch Repair: Mechanisms and Relationship to Cancer Susceptibility. Trends in Biochemical Sciences 1995;20(10):397-401.

11. Liu, Z., Wang, L. E., Strom, S. S., Spitz, M. R., Babaian, R. J., DiGiovanni, J., and Wei, Q. Overexpression of HMTH in Peripheral Lymphocytes and Risk of Prostate Cancer: a Case-Control Analysis. Mol.Carcinog. 2003;36(3):123-9.

12. Shen, M. R., Jones, I. M., and Mohrenweiser, H. Nonconservative Amino Acid Substitution Variants Exist at Polymorphic Frequency in DNA Repair Genes in Healthy Humans. Cancer Res. 2-15-1998;58(4):604-8.

13. De Marzo, A. M., Marchi, V. L., Epstein, J. I., and Nelson, W. G. Proliferative Inflammatory Atrophy of the Prostate: Implications for Prostatic Carcinogenesis. Am.J.Pathol. 1999;155(6):1985-92.

14. Nelson, W. G., De Marzo, A. M., DeWeese, T. L., Lin, X., Brooks, J. D., Putzi, M. J., Nelson, C. P., Groopman, J. D., and Kensler, T. W. Preneoplastic Prostate Lesions: an Opportunity for Prostate Cancer Prevention. Ann.N.Y.Acad.Sci. 2001;952:135-44.

15. Giovannucci, E., Rimm, E. B., Liu, Y., Stampfer, M. J., and Willett, W. C. A Prospective

Study of Tomato Products, Lycopene, and Prostate Cancer Risk. J.Natl.Cancer Inst. 3-6-2002;94(5):391-8.

16. Xu, J., Zheng, S. L., Turner, A., Isaacs, S. D., Wiley, K. E., Hawkins, G. A., Chang, B. L., Bleecker, E. R., Walsh, P. C., Meyers, D. A., and Isaacs, W. B. Associations Between HOGG1 Sequence Variants and Prostate Cancer Susceptibility. Cancer Res. 4-15-2002;62(8):2253-7.

17. Goode, E. L., Ulrich, C. M., and Potter, J. D. Polymorphisms in DNA Repair Genes and Associations With Cancer Risk. Cancer Epidemiol.Biomarkers Prev. 2002;11(12):1513-30.

18. Kassie, F., Parzefall, W., and Knasmuller, S. Single Cell Gel Electrophoresis Assay: a New Technique for Human Biomonitoring Studies. Mutat.Res 2000;463(1):13-31.

19. Collins, A. R., Dusinska, M., Horvathova, E., Munro, E., Savio, M., and Stetina, R. Inter-Individual Differences in Repair of DNA Base Oxidation, Measured in Vitro With the Comet Assay. Mutagenesis 2001;16(4):297-301.

20. Berwick, M. and Vineis, P. Markers of DNA Repair and Susceptibility to Cancer in Humans: an Epidemiologic Review. J Natl Cancer Inst. 6-7-2000;92(11):874-97.

21. Rajaee-Behbahani, N., Schmezer, P., Risch, A., Rittgen, W., Kayser, K. W., Dienemann, H., Schulz, V., Drings, P., Thiel, S., and Bartsch, H. Altered DNA Repair Capacity and Bleomycin Sensitivity As Risk Markers for Non-Small Cell Lung Cancer. Int.J Cancer 3-20-2001;95(2):86-91.

22. Saha, D. T., Davidson, B. J., Wang, A., Pollock, A. J., Orden, R. A., and Goldman, R. Quantification of DNA Repair Capacity in Whole Blood of Patients With Head and Neck Cancer and Healthy Donors by Comet Assay. Mutat.Res. 1-31-2008;650(1):55-62.

23. Singh, N. P., McCoy, M. T., Tice, R. R., and Schneider, E. L. A Simple Technique for Quantitation of Low Levels of DNA Damage in Individual Cells. Exp.Cell Res 1988;175(1):184-91.

24. Schmezer, P., Rajaee-Behbahani, N., Risch, A., Thiel, S., Rittgen, W., Drings, P., Dienemann, H., Kayser, K. W., Schulz, V., and Bartsch, H. Rapid Screening Assay for Mutagen Sensitivity and DNA Repair Capacity in Human Peripheral Blood Lymphocytes. Mutagenesis 2001;16(1):25-30.

25. Zheng, Y. L., Loffredo, C. A., Yu, Z., Jones, R. T., Krasna, M. J., Alberg, A. J., Yung, R., Perlmutter, D., Enewold, L., Harris, C. C., and Shields, P. G. Bleomycin-Induced Chromosome Breaks As a Risk Marker for Lung Cancer: a Case-Control Study With Population and Hospital Controls. Carcinogenesis 2003;24(2):269-74.

26. Xu, X., Veenstra, T. D., Fox, S. D., Roman, J. M., Issaq, H. J., Falk, R., Saavedra, J. E., Keefer, L. K., and Ziegler, R. G. Measuring Fifteen Endogenous Estrogens Simultaneously in Human Urine by High-Performance Liquid Chromatography-Mass Spectrometry. Anal.Chem. 10-15-2005;77(20):6646-54.

27. Xu, X., Roman, J. M., Veenstra, T. D., Van Anda, J., Ziegler, R. G., and Issaq, H. J. Analysis of Fifteen Estrogen Metabolites Using Packed Column Supercritical Fluid Chromatography-Mass Spectrometry. Anal.Chem. 3-1-2006;78(5):1553-8.

KEY RESEARCH ACCOMPLISHMENTS:

1. We have enrolled 131 cases and 231 controls (25% African American participants) into our case-control study of prostate cancer. A fully annotated repository of serum, plasma, white blood cells, and urine was established.
2. The comet assay optimization was completed. We developed a procedure for quantification of DNA repair capacity (22) and applied the assay to our case-control population. The results show that the comet DNA repair phenotype is nort associated with case-control status. However, we observed a slower DNA repair capacity in African American controls compared to Caucasian controls. This might be associated with higher risk of prostate cancer in the African American men and should be further examined.
3. Mutagen sensitivity assay following chromatid breaks in short term cultured white blood cells exposed to bleomycin showed significant cases-control differences in a subset of participants. We did not have a sufficient number of African Americans in theis subset to evaluate the racial differences in risk.
4. We have expanded the analyses to an analysis of 14 metabolites of estrogen in urine of the paqrticipants in order to evaluate the effects of estrogenic signaling on prostate cancer risk. Preliminary analysis did not detect any case-control differences; however, more detailed analyses are needed and under way.

REPORTABLE OUTCOMES:

We have published three manuscripts in course of this study the acknowledge the support of the Department of Defense grant W81XWH-04-1-0294 awarded to RG. Two additional manuscripts describing the comet study and the estrogen metabolite study are submitted or in preparation (see appendix).

1. Saha, D. T., Davidson, B. J., Wang, A., Pollock, A. J., Orden, R. A., and **Goldman, R**. Quantification of DNA Repair Capacity in Whole Blood of Patients With Head and Neck Cancer and Healthy Donors by Comet Assay. Mutat.Res. 1-31-2008;650(1):55-62.

2. **Goldman, R.**, Ressom, H. W., Abdel-Hamid, M., Goldman, L., Wang, A., Varghese, R. S., An, Y., Loffredo, C. A., Drake, S. K., Eissa, S. A., Gouda, I., Ezzat, S., and Moiseiwitsch, F. S. Candidate Markers for the Detection of Hepatocellular Carcinoma in Low-Molecular Weight Fraction of Serum. Carcinogenesis 2007;28(10):2149-53.

3. Ressom, H. W., Varghese, R. S., Drake, S. K., Hortin, G. L., Abdel-Hamid, M., Loffredo, C. A., and **Goldman, R**. Peak Selection From MALDI-TOF Mass Spectra Using Ant Colony Optimization. Bioinformatics. 3-1-2007;23(5):619-26

Several posters were presented at scientific meetings.

A poster was presented at the 94th annual meeting of the American Association for Cancer Research (AACR) in April 2006 in Washington, DC and at the Annual LCCC Research Competition, Georgetown University, February 2007.

Aleksandra Dakic, Allison Pollock, Michelle Ma, Daniel Saha, Sara Samie, Sherine Salem, Bozena Novotna, and **Radoslav Goldman**. Optimization of Comet assay for quantification of DNA repair capacity in human whole blood. 97th Annual AACR Conference, Washington, DC, April 2006

Daniel Saha, Tony Orden, Bozena Novotna, and **Radoslav Goldman**. Use of Comet Assay for Quantification of DNA Repair Capacity in Human Whole Blood in a Prostate Cancer study. Annual LCCC Research Competition, Georgetown University, February 2007

CONCLUSIONS:

We have overcome the initial slow recruitment and completed a case control study examining DNA repair differences in prostate cancer risk of African American and Caucasian men. We did not observe a direct association of the comet DNA repair phenotype with cancer risk. We did, however, observe a slower DNA repair phenotype in African American men.

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

Informed Consent for Clinical Research

MedStar Research Institute/Georgetown Medical Center

INSTITUTION: GUMC + WHC

INTRODUCTION

We invite you to take part in a research study. The study is called 'Molecular Epidemiology of Prostate Cancer.' Please take your time to make your decision. Discuss it with your family and friends. It is important that you read and understand several general principles that apply to all who take part in our studies:

(a) Taking part in the study is entirely voluntary;

(b) Personal benefit to you may or may not result from taking part in the study, but knowledge may be gained from your participation that will benefit others;

(c) You may withdraw from the study at any time without any of the benefits you would have received normally being limited or taken away.

The nature of the study, the benefits, risks, discomforts and other information about the study is discussed below. Any new information discovered, at any place during the research, which might affect your decision to participate or remain in the study will be provided to you. You are urged to ask the staff members any questions you have about this study and the staff members will explain the questions to you. The investigator (person in charge of this research study) is Dr. Radoslav Goldman. The research is being sponsored by the Department of Defense. The Department of Defense is called the sponsor and the Georgetown University is being paid by the Department of Defense to conduct this study with Dr. Radoslav Goldman as the primary investigator.

WHY IS THE STUDY BEING DONE?

Study participants include cases and controls.

If you are a **case**, you are being asked to participate in this study because you are suspected of having prostate cancer or have prostate cancer. Your prostate tumor, blood and other samples may show us how cancer develops and what are the factors that helped increase the cancer risk.

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

If you are a **control**, you are being asked to participate in this study because a comparison group free of prostate cancer is needed to evaluate the results. Your blood and other samples may show us how cancer develops and what the factors are that help increase cancer risk.

The purpose of this study is to learn about the natural history of prostate cancer and its causes and treatments. This research is being done because the causes of prostate cancer are not well understood at present. The purpose of this research is to see how someone's ability to respond to genetic damage modifies risk of prostate cancer. We will test how your ability to repair damaged DNA and eliminate cells that did not repair the damage modifies prostate cancer risk.

We will examine your blood, cheek swabs, saliva, nail clippings and urine to see if tests for your response to chemical exposure can help us predict who might be at greater risk of prostate cancer. If you are going to have surgery, or had surgery, or if you are going to have a biopsy or had a biopsy, we will use samples of tumor tissue, as well as adjacent normal tissue, to determine whether markers in the tissue suggest how the cancer developed. The specimens will not be used for diagnostic purposes or for purposes related to your medical care. That is, the experiments done on these samples will not be used for decisions about your personal risk of prostate cancer, your treatment or your prognosis. These specimens will be available to qualified medical researchers for scientific studies that have been approved by the Principal Investigator, listed above, and an oversight committee. Researchers who receive these samples will not have access to your name or other identification information.

Cases: If you wish, you will be given the opportunity to identify friends living in your geographical area to be controls in the study. This would help us to identify a group of controls subjects without prostate cancer. We hope that this research can lead to the discovery of new tests for cancer risk, including genetic tests.

Men older than 18 years of age free of prostate cancer are eligible to participate as **controls** in this study. To minimize the possibility that you have undetected prostate cancer, we will perform a test for prostate specific antigen (PSA) on a portion of your blood sample free of charge to you. If your test shows a PSA value greater than 2.5ng/ml, a follow up examination by a doctor will be recommended.

All men at all stages of presentation are eligible to participate as **cases** in this study.

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

HOW MANY PEOPLE WILL TAKE PART IN THE STUDY?

About 600 people (300 patients and 300 controls) will take part in this study and will be recruited at Washington Hospital Center and Georgetown University Medical Center. Participants in the study are referred to as "subjects".

WHAT IS INVOLVED IN THE STUDY?

Upon reviewing and signing this informed consent, you will begin the study. We will ask you questions using a form that will take about an hour to finish. If you do not want to do the whole questionnaire at the time you give blood, we can do only one part lasting about 15 minutes and then we will contact you later to finish the study. Your blood, cheek cells, saliva, nail tissue, and urine will be tested for their response to chemical exposure, in order to identify tests that may predict cancer risk. This research will be conducted on an experimental basis only, and you will not be provided with any information about your test results.

If you take part in this study, you will have the following tests and procedures:

1. Upon reviewing and signing this informed consent, you will begin the study.
2. Undergo an in-person interview lasting about one hour administered by a trained interviewer.
3. Provide a blood sample that is about 3 tablespoons.
4. Provide a urine specimen.
5. Provide two cheek swab samples.
6. Provide saliva.
7. Provide nail clippings.
8. Complete and return a self-administered diet history questionnaire.

Additionally, **cases** will:

9. Allow us to use the unneeded portion of your prostate tissue, as well as a small sample of adjacent normal tissue for research purposes.

HOW LONG WILL I BE IN THE STUDY?

We expect that your participation in the study will take about an hour in addition to any scheduled examination. The study is completed after you finish your questionnaires and donate your blood, urine, nail clippings, saliva, a cheek sample and for **cases** only, tissue from surgery/biopsy not needed for diagnostic purposes. However, if you agree below, we may call you in the future for additional information and/or sample collection. We will use your sample for different tests as described above and as new hypotheses develop for as long as it lasts and is useful for our testing. If the sample is no longer useful, it will be destroyed. However, you can request that your blood, cheek cells, saliva, nail tissue, urine and prostate tissues be destroyed at any time. To have your samples destroyed, you can contact Dr. Goldman at 202-687-9868.

The investigators, physicians or sponsors may stop the study or take you out of the study at any time should they judge that it is in your best interest to do so, if you experience a study-related injury, or if you do not comply with the study plan. They may remove you from the study for various other administrative and medical reasons. They can do this without your consent.

In the future, it might be necessary to contact you for further information or an additional blood sample (or other type of biological sample). If this is okay, please indicate below. You can refuse to do so now or later. Please check and initial below:

I ____may ____may not be contacted in the future for further information or biological samples.

_____ Sign your initials here.

WHAT ARE THE RISKS OF THE STUDY?

There is a very slight chance of a bruise or an infection from the blood draw, but we use only trained medical technicians to draw your blood and they will use the best available precautions. Another possible risk is that your genetic information might be obtained by persons outside the study. We will minimize this chance by maintaining the confidentiality of your test results and study records at all times (see below). For more information about risks and side effects, ask the research staff or contact Radoslav Goldman at 202-687 9868.

ARE THERE ANY BENEFITS TO TAKING PART IN THE STUDY?

If you agree to take part in this study, there is no direct medical benefit to you. We hope the information learned from this study will benefit others in the future.

WHAT ABOUT CONFIDENTIALITY?

Efforts will be made to protect your personal information to the extent allowed by law. Medical records of research study participants are stored and kept according to legal requirements. You will not be identified in any reports or publications resulting from this study. Organizations that may request, inspect and/or copy your research and medical records for quality assurance and data analysis include groups such as: Department of Defense, Food and Drug Administration, MedStar Research Institute, Georgetown University, and Institutional Review Board (IRB).

We will store your tissue, blood, cheek, saliva, nail and urine samples, or genetic material prepared from your blood, urine, cheek, saliva, nail or prostate tissue, in a secure room with restricted access. Only people working on this research project can work on your samples. Because we want to protect your confidentiality, your samples will have only a number on the tube and will not have your name or other identifier information.

We will protect your genetic and other testing results. We will control access to the computer files that hold this information. Access to the computer files can only be obtained through multiple passwords. Only authorized study personnel can link your sample to you. This information will not be released to anyone. "Anyone" includes you, your family, your doctor, your insurance company, or your employer. This is because the research is at a very early stage and we would not be able to tell you what your results mean. This information will not be included in any medical records.

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

CERTIFICATE OF CONFIDENTIALITY

To help us protect your privacy, we have obtained a Certificate of Confidentiality from the National Institutes of Health. With this Certificate, the researchers cannot be forced to disclose information that may identify you, even by a court subpoena, in any federal, state, or local civil, criminal, administrative, legislative, or other proceedings. The researchers will use the Certificate to resist any demands for information that would identify you, except as explained below.

The Certificate cannot be used to resist a demand for information from personnel of the United States Government that is used for auditing or evaluation of Federally funded projects or for information that must be disclosed in order to meet the requirements of the federal Food and Drug Administration (FDA).

You should understand that the Certificate of Confidentiality does not prevent you or a member of your family from voluntarily releasing information about yourself or your involvement in this research. If an insurer, employer, or other person obtains your written consent to receive research information, then the researchers may not use the Certificate to withhold that information.

WHAT ARE THE COSTS?.

There is no cost to participate in the study

You should not expect any one to pay you for pain, worry, lost income, or non-medical care costs that occur from taking part in this research study.

You or your insurance company will be charged for continuing medical care and/or hospitalization that are not a part of the study.

RESEARCH RELATED INJURY

The Department of Defense is partially funding this research. Should you be injured as a direct result of participating in this research, you will be provided medical care at no cost to you. You will not receive any injury compensation, only medical care. Your insurance company will be billed, but you will not be liable for any costs not covered by your insurance. Additional information on this subject may be obtained from the Office of the Medical Director, Georgetown University Hospital at (202) 784-3011.

You *will not* be paid for participating in this study.

COMMERCIAL INTEREST

On rare occasions, laboratory research on human specimens results in discoveries that are the basis for new research products or diagnostic and therapeutic methods. It is the policy of Georgetown University Medical Center, MedStar, Inc., and their affiliates not to compensate you for any future financial claim to your tissues for research and development for commercial and noncommercial purposes**.** No funds are available or will be paid by the MedStar Research Institute, MedStar Health or Georgetown University to repay you in case of injury.

I understand that I will not receive financial compensation for my biological samples at any time. _____(sign initials here)

WHAT ARE MY RIGHTS AS A PARTICIPANT?

Taking part in this study is voluntary. You may choose not to take part in or leave the study at any time. If you request, the link between your name and the study results will be destroyed. Also, your biological samples will be discarded at your request. However, the results of any finished analysis and or published result will be kept to preserve the validity of the study. If you choose to not take part in or to leave the study, your regular care will not be affected and you will not lose any of the benefits you would have received normally.

We will tell you about new information that may affect your health, welfare, or participation in this study.

We will not provide you with any of the results we obtain from your biological samples.

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

WHO DO I CALL IF I HAVE QUESTIONS OR PROBLEMS?

For questions about the study, problems, unexpected physical or psychological discomforts or injuries related to the study, contact day or night the research doctor, Radoslav Goldman at 202-687-9868. If you would like to write to him, please send mail to: Radoslav Goldman, Georgetown University, 3800 Reservoir Road NW, Lower Level S-183, Washington DC 20057.

If you are a participant at Washington Hospital Center and have questions about your rights as a research participant, contact the MedStar Research Institute. Direct your questions to Dr. Barbara Howard at Medstar Research Institute:

MedStar Research Institute
6495 New Hampshire Ave., Suite 201
Hyattsville, MD 20783
Tel: (301) 853-7532
Pager: 1-888-663-6842

Or

If you are a participant at Georgetown University Medical Center and have questions about your rights as a research participant, contact the Georgetown University IRB Office. Direct your questions to:

Ms. Laura Miller, Executive Officer, Institutional Review Board at:
Address: Georgetown University Medical Center
3900 Reservoir Road, N.W.
NE 105 Med-Dent
Washington, D.C. 20007
Telephone: (202) 687-1506

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

SIGNATURES

As a representative of this study, I have explained the purpose, the procedures, the benefits and risks that are involved in this research study. Any questions that have been raised have been answered to the individual's satisfaction.

_____ _____

Signature of person obtaining the consent Date

I, the undersigned have been informed about this study's purpose, procedures, possible benefits and risks, and I have received a copy of this consent. I have been given the opportunity to ask questions before I sign, and I have been told that I can ask other questions at any time. I voluntarily agree to participate in this study. I am free to withdraw from the study at any time without need to justify my decision. This withdrawal will not in any way effect my future treatment or medical management. I agree to cooperate with Dr. Radoslav Goldman and the research staff and to inform them immediately if I experience any unexpected or unusual symptoms.

_____

Printed name of subject

_____

Printed permanent address of subject.

_____ _____

Signature of Subject Date

_____ _____

Signature of Witness Date

_____ _____

Principal Investigator (if not person obtaining consent) Date

Study number: Principal Investigator (s): Radoslav Goldman
Title Molecular Epidemiology of Prostate Cancer

FOLLOW-UP SAMPLE ACQUISITION CONSENT

As a continuation of the study in which I enrolled on _____ (date), I agree to provide a set of biological samples including urine, blood (about 3 tablespoons), cheek cells, and saliva and to answer questions about my medical history. In case I undergo surgery to remove a tumor, I agree to donate the unneeded portion of my prostate tissue as well as adjacent normal tissue removed at surgery for research purposes. I, the undersigned, have been informed about this study's purpose, procedures, possible benefits and risks, and I have received a copy of this consent. I have been given the opportunity to ask questions before I sign, and I have been told that I can ask other questions at any time. I voluntarily agree to participate in this study. I am free to withdraw from the study at any time without need to justify my decision. This withdrawal will not in any way effect my future treatment or medical management. I agree to cooperate with Dr. Radoslav Goldman and the research staff and to inform them immediately if I experience any unexpected or unusual symptoms related to the research study.

_____ _____

Signature of Subject Date

_____ _____

Signature of Witness Date

_____ _____

Principal Investigator (if not person obtaining consent) Date

MedStar Research Institute-Georgetown University Oncology Institutional Review Board Application (Protocol) IRB Review (AB-1)

Section One: Application Information

Principal Investigator	Radoslav Goldman, Ph.D.
Department	Oncology
Title	Assistant Professor
Phone/Pager: 202-687 9868	**Fax:** 202-687 1988
E-mail address:rg26@georgetown.edu	
Mailing Address: Georgetown University, Lombardi Cancer Center, LL (S) Level, Room 183, 3800 Reservoir Rd. NW, Washington DC 20057	
Co-Investigator: Christopher Loffredo, Department of Oncology	
Title: Assistant Professor	
Phone/Pager: 202-6873758	**Fax:** 202-7843034
Email address: cal9@georgetown.edu	
Mailing Address: Georgetown University, S-153, 3800 Reservoir Rd. NW, Washington DC 20057	
Study Coordinator (member of faculty or administrative official) Alexandra Schopf	

Title of Project	**Purpose of Project (one or two sentences)**
Molecular Epidemiology of Prostate Cancer	This study has two goals: 1. To establish a prostate cancer data and tissue repository; and 2. To utilize the repository to test whether prostate cancer is related to interindividual variability in the response to genotoxic stress.

Consultants, if any	**Department or Institution**
Asim Amin, M.D.	Medicine and Oncology, Georgetown University
Anatoly Dritschilo, M.D.	Radiation Medicine, Georgetown University
John Lynch, M.D.	Urology, Georgetown University
John Lynch, M.D.	Urology, Georgetown University
Peter Shields, M.D.	Oncology, Georgetown University
Bhaskar Kalakouri, M.D.	Pathology, Georgetown University
Mohan Verghese, M.D.	Radiation Oncology, Washington Hospital Center
Michael Porrazzo, M.D.	Urologic Oncology, Washington Hospital Center
Pamela Randolph, M.D.	Medical Oncology, Washington Hospital Center

Estimated duration of total project	3 years
Estimated total number of subjects (including control subjects)	600
Age range of subjects	>18

Sex of subjects	Male
Where will study be conducted?	GUMC
Source of subjects	Georgetown University Hospital and Washington Hospital Center

Grant Support for Project (if any)	**Commercial Support (if any) for Project**
Funded in part by the Department of Defense. Additional funding will be provided by the Lombardi Cancer Center and the protocol will be conducted by the GCRC laboratory. Once pilot data is obtained, additional grant funding will be sought.	

Investigational New Drug (IND)	**Investigational Device Exemptions (IDE)**
❑ None ❑ IND: FDA No. _____ ❑ Drug Name: _____ ❑ Drug Sponsor: _____	❑ None ❑ IDE: FDA No. _____ ❑ Device Name: _____ ❑ Device Sponsor: _____ ❑ Significant (SR) ❑ Non-Significant Risk (NSR)

Section Two: Additional MedStar Research Institute-Georgetown University Regulatory Information

1. Does this project involve the use of biohazardous materials, recombinant DNA and/or gene therapy?
   - ❑ Yes. If so, Institutional Biosafety Committee (IBC) approval must be obtained. Contact 202-687-4712 for assistance.
   - √ No.

2. Has the Institutional Biosafety Committee approved the protocol?

√ NA

❑ **Approved**	**Date Approved:**
❑ **Application Pending**	**Date Submitted:**

3. Does this project include the use of radioisotopes and/or radiation-producing devices regardless of whether the use is incidental to the project?
   - ❑ Yes. If so, all protocols must be submitted to the GUH RSC along with a completed RSC-4 or RSC-5 form. The forms require information on the use of radioisotopes and radiation-producing devices and must include dose calculations. Call 202-687-4712 to obtain forms or if additional information is required.
   - ❑ No.

4. Has the Radiation Safety Committee approved the protocol?

√ NA

❑ **Approved**	**Date Approved:**
❑ **Application Pending**	**Date Submitted:**

5. Does this project involve the use of fetal tissue?
   - ❑ Yes
   - √ No

6. Do any investigators or co-investigators have a conflict of interest as defined in the Georgetown University Faculty handbook or MedStar Health Institute policy?
   - ❑ Yes. If yes, please explain.
   - √ No

7. A copy of each investigator's current Conflicts of Interest Disclosure Form must be attached to this application.

If this project involves a FDA regulated drug or device, you must file a FDA form 3455.

Section Three: Information for Protocol Review
Please answer each specific question and use additional sheets as needed. A response of "See attached protocol or grant application" is not sufficient.

6. Provide a brief historical background of the project with reference to the investigator's personal experience and to pertinent medical literature. Use additional sheets as needed.

Despite the fact that prostate cancer is the most common tumor among US males, relatively little is known about the causative mechanisms. The known risk factors include age, ethnicity or race, high-fat diet and family history of prostate cancer, but these factors are not sufficient for identification of men with increased susceptibility. Establishing new biomarkers of cancer risk would greatly benefit the field of prostate cancer prevention and surveillance.

Mutagen sensitivity and comet assay are established biomarkers of risk (1). The mutagen sensitivity assay measures response to a genotoxic insult (e.g. bleomycin exposure) in short-term cultured human lymphocytes in terms of the number of chromatid breaks; comet assay measures DNA unwinding under alkaline conditions. Subjects with a high number of chromatid breaks in mutagen sensitivity assay or high DNA unwinding in comet assay have higher cancer risk. For example, comparison of cancer risk in the highest/lowest quartile of mutagen sensitivity in a study of 150 head and neck cancer cases and 150 controls matched on age and race showed an odds ratio of 4.5 with p=0.04 (2). Surprisingly, these phenotypic assays were not yet examined in prostate cancer. Even though the exact mechanism underlying the phenotypes is unknown, variability in DNA-repair capacity is consistent with the available experimental results (3). Moreover, it was shown in twin studies that mutagen sensitivity is heritable in non-cancer subjects. The correlation coefficient was 0.79 (95% confidence interval = 0.65-0.88) in monozygotic twins while for dizygotic twins the coefficient was 0.42 (95% confidence interval = 0.00-0.71) (4). Mutagen sensitivity and comet assay phenotypes therefore reflect multiple genetic traits related to DNA repair capacity, which predispose an individual to cancer risk.

Apoptosis is a molecular pathway eliminating, besides other functions, cells unable to cope efficiently with genotoxic stress. Deficient apoptosis is a likely candidate for a cancer-prone phenotype. Apoptosis was implicated in regulation of response to radiation therapy in prostate cancer (5), malignancy of prostatic tumor (6), and recurrence of prostate carcinoma following surgery (7). For example, in 54 prostate cancer patients treated with radiotherapy the response was negative in 84% cases with positive bcl-2 immunohistochemistry and bcl-2 was an independent prognostic variable for treatment with odds ratio of 7.3 (5). Apoptotic index was associated with disease recurrence in a study of 47 men following radical prostatectomy (7). But apoptosis was not yet examined as a phenotypic predictor of prostate cancer risk. Since the apoptotic phenotype is a composite measure of a number of converging mechanistic pathways, it is advantageous to the measurement of each individual genotype in the pathway.

Lipid peroxidation was suggested as a mechanism underlying the association of dietary fat and prostate cancer risk. Lipid peroxidation leads to oxidative genotoxic stress, that can overwhelm DNA repair and/or apoptotic mechanisms and potentially lead to cancer. We propose to quantify malondialdehyde deoxyguanosine adducts (dGMDA) in peripheral blood lymphocytes and prostate tumors. HPLC methods will be used for all assays.

DNA repair consists of two major categories, excision repair (base excision repair and nucleotide excision repair) and recombination repair (homologous and non-homologous) (8). Numerous polymorphisms in the DNA repair genes have been identified (9) and are likely to contribute to cancer risk through decreased efficiency of response to genotoxic stress. But two functional polymorphisms in DNA repair genes, *OGG1* and *XRCC1*, are particularly relevant to this study. Both genes are involved in the repair of 8-hydroxy-guanine (8-OHdG) and other oxidative lesions (10); and our study examines mainly how variability in the response to oxidative DNA damage modifies risk for prostate cancer

(bleomycin is a radiomimetic which induces oxidative DNA damage and mutagen sensitivity is mainly a model of this pathway). OGG1 is a DNA glycosylase/AP lyase involved in base excision repair of 8-OHdG and XRCC1 is a DNA ligase III terminating the base excision repair cascade (10). The OGG1 Ser(321)Cys polymorphism codes for a protein with a lower 8-OHdG repair capacity and leads to several splicing variants of unknown functional significance (11). This variant occurs at a frequency of 0.4 in Japanese and was associated with an increased risk of lung cancer in a study of 241 cases and 197 controls with an OR=3.01 (95% CI 1.33-6.83) (12). This variant was found in a Caucasian population at a frequency of 0.22 and was not associated with lung cancer in this study (13). Examination of this polymorphism in prostate cancer is therefore highly relevant. The XRCC1 Arg(399)Gln polymorphism was associated with increased sensitivity of human lymphocytes to DNA damage (14), increased risk of squamous cell carcinoma of the head and neck (15), increased risk of early onset colorectal carcinoma (16), and increased risk of adenocarcinoma of the lung (17). The polymorphism occurs in 37% of Caucasians and 17% of African-Americans (19). An examination of the XRCC1 'at risk' polymorphism as a risk factor for prostate cancer was not reported.

The study of mutations in human tumors and experimental models is elucidating important carcinogenic mechanisms (20). The study of mutations in the p53 tumor suppressor gene is uniquely suited for the study of cancer etiology, because p53 is involved in many cellular processes (including maintenance of genomic stability, programmed cell death, and DNA repair) and in tumors often accumulates point mutations amenable to further analysis (21). Specific mutations in p53 can reflect carcinogenic insults that precede cancer. It was shown that reactive oxygen species are a major source of G:C -> A:T transitions at non-CpG sites. For example, in radiation-induced lung cancer, G:C -> A:T transitions at non-CpG sites dominate the p53 mutational spectra, which differs markedly from mutational spectra associated with tobacco (22,23). Oxidatice damage is expected to be a major source of DNA damage in prostate cancer. Mutagen sensitivity and comet assay are a model of oxidative DNA damage (bleomycin is a radiomimetic which induces oxidative DNA damage), and *OGG1* and *XRCC1* participate in the repair of oxidatively damaged DNA. We therefore predict that G:C -> A:T transitions at non-CpG sites will correlate with mutagen sensitivity/comet assay phenotypes and at risk variants of *OGG1* and *XRCC1*. This study would provide for the first time an evidence for such an association. The p53 gene is also an attractive target because it is mutated in up to 35% of early prostate cancers (24).

Significance: We are proposing a molecular epidemiology study to test variation in the response to genotoxic stress and in DNA repair as a biomarker of prostate cancer risk. This study measures mutagen sensitivity, comet assay, apoptosis, and polymorphism in *OGG1* and *XRCC1* as biomarkers of prostate cancer risk; the study also correlates mutations in p53 tumor supressor gene with mutagen sensitivity. The proposal is innovative because neither of the proposed biomarkers was to our knowledge examined in connection with prostate cancer risk. If mutagen sensitivity, apoptosis, or DNA repair-variants correlate with prostate cancer risk, they could serve as readily obtainable biomarkers to identify men with increased risk of prostate cancer. The phenotypic biomarkers could be used to better identify the currently poorly understood genotoxic insults leading to cancer risk (improved risk models in case-control studies). Elucidating mechanisms of the early stages of prostate carcinogenesis would have an immediate impact for prevention and surveillance. Better prevention strategies (including chemoprevention) could be designed and tested based on the identified targets. And new hypotheses focusing on the genetic and environmental factors associated with prostate cancer risk could be formulated and evaluated.

Dr. Radoslav Goldman, Principal Investigator*:* Dr. Goldman is Assistant Professor of Oncology and a member of the Cancer Genetics and Epidemiology Program at LCC. He is an analytical toxicologist with specialization in biomarker studies of cancer risk. Dr. Goldman will be responsible for the design and execution of the proposed study, data analysis, and result interpretation. He will work in close collaboration with Dr. Loffredo and Dr. Shields on the establishment of the prostate biomarker resource.

Dr. Christopher Loffredo, Co-Investigator*:* Dr. Loffredo is Assistant Professor of Oncology and a member of the Cancer Genetics and Epidemiology Program at LCC. He is responsible for the

epidemiological field activities of the Biomarker Core Resource. Dr. Loffredo will assist with the coordination of the collection and transfer of specimen, repository, and statistical analyses.
Dr. Asim Amin, Consultant: Dr. Amin is Assistant Professor of Medicine and Oncology. He will refer patients from this department to the study coordinator.
Dr. Anatoly Dritschilo, Consultant: Dr. Dritschilo is Professor and Chairman of the Department of Radiation Oncology and will refer patients from this department to the study coordinator.
Dr. John Lynch, Consultant: Dr. Lynch is Professor of Surgery and Chairman of the Department of Urology. He will refer patients from this department to the study coordinator.
Dr. Peter Shields, Consultant: Dr. Shields is Professor of Oncology and Medicine, Director of Cancer Genetics and Epidemiology Division, and Associate Director for Population Sciences. Dr. Shields will assist in the design and oversight of the study.
Dr. Bhaskar Kalakouri , Consultant: Dr. Singh is Assistant Professor of Pathology and will oversee the collection and processing of prostate tissue for this study.
Dr. David Perry, Consultant: Dr. Perry is Medical Director of Clinical Research, Washington Hospital Center, and will refer patients to the study and help us coordinate recruitment effort at this hospital.
Dr. Mohan Verghese, Consultant: Dr. Verghese is from the Department of Radiation Oncology, Washington Hospital Center, and will refer patients from this department to the study coordinator.
Dr. Michael Porrazzo, Consultant: Dr. Porrazzo is from the Department of Urologic Oncology, Washington Hospital Center, and will refer patients from this department to the study coordinator.
Dr. Pamela Randolph, Consultant: Dr. Randolph is from the Department of Medical Oncology, Washington Hospital Center, and will refer patients from this department to the study coordinator.

7. The plan of study. State the hypothesis or research question you intend to answer. Describe the research design and procedures (including standard procedures) to be used in the research. Specifically identify any experimental procedures. Provide statistical justification for the number of subjects to be studied and the degree of change expected. Describe any special equipment or unusual procedures to be used for this research project. Use additional sheets as needed.

Research Question: This study has two goals: 1. To establish a prostate cancer data and tissue repository; and 2. To utilize the repository to test our hypothesis that prostate cancer is related to interindividual variability in the response to genotoxic stress. We propose to examine 1. Mutagen sensitivity, comet assay, and apoptotic response to bleomycin in peripheral blood lymphocytes; 2.; dGMDA adduct in lymphocytes and prostate tissue and 3. Genetic variants of the DNA repair genes *OGG1* and *XRCC1* as biomarkers of prostate cancer risk. In selected cases, we will examine the association of p53 mutational spectrum with mutagen sensitivity and genetic polymorphisms in *XRCC1* and *OGG1*.

Specific Aims: This study can address several areas of prostate cancer by developing the infrastructure to allow us to identify new biomarkers of prostate cancer risk, and improve our ability to optimize prevention and treatment strategies for prostate cancer. We plan to develop an ongoing recruitment of prostate cancer cases so that we can study prostate tumor tissue, blood and other specimen in order to understand the genotypic and phenotypic expression (e.g., mutagen sensitivity) of possible prostate cancer risk markers and to establish genotype-phenotype relationships. By linking an epidemiological profile to the tissue tumor markers, we will be able to elucidate gene-environment interactions by performing a case-control analysis and searching for etiological clues in the tumor tissue (e.g. p53 mutational spectra). The genetic risk markers under study will be limited to low penetrance genes that modulate the risk of prostate cancer and carry a risk in the context of prostate cancer of about 2-fold.

The specific aims and hypotheses of this project are to:

1. Recruit prostate cancer cases and controls to provide an epidemiological profile, blood, urine, nail clipping, and tumor tissue (when available). This will establish a data and tissue repository.

2. Utilize the repository to study low penetrance genes, investigate gene-environment interactions and establish genotype-phenotype relationships involving DNA damage, DNA repair and response to DNA damage, in order to identify or validate the use of intermediate biomarkers of cancer risk.

H_{2a} High mutagen sensitivity/comet assay increase the risk of prostate cancer.
H_{2b} Low apoptotic response increases the prostate cancer risk.
H_{2c} High dGMDA adducts increase prostate cancer risk.
H_{2d} At risk variants of XRCC1 and OGG1 increase prostate cancer risk.

3. To identify the relationship of biomarkers measured in surrogate tissues such as blood, buccal swabs and urine to pathological markers in prostate tumor. Investigate gene-environment interactions and establish genotype-phenotype relationships involving DNA damage, and response to DNA damage, in order to identify or validate the use of intermediate biomarkers of cancer risk.

H_{3a} Comet assay/dGMDA in lymphocytes correlate with these markers in prostate tissue.
H_{3b} Genetic polymorphism of DNA repair-genes is associated with p53 mutations.
H_{3c} Mutagen sensitivity is associated with p53 mutations.

Methods: Cases will be enrolled from the Departments of Medicine and Oncology, Radiation Medicine, and Urology at the Georgetown University Medical Center and Washington Hospital Center.

Approximately 200 newly diagnosed patients with prostate cancer are treated currently each year at each clinic, which is more then enough for our goal to enroll 300 patients in three years. All participants will be requested to complete an informed consent and undergo a forty five minute interview, phlebotomy, buccal cell collection and provide a nail clipping and urine sample. Also unneeded pathological tissue from patients (tumor and adjacent normal tissue) will be collected if available. A repository will be established for future studies as new hypotheses are generated.

The weekly schedule for the clinic is available to the phlebotomist/interviewer so that he/she can determine the times when eligible patients are in the clinic. Most such patients are seen at the clinic once or twice prior to their surgery so there is ample opportunity to enroll them prior to any treatment. Dr. Amin and the other consultants will inform the patients about the study and those who are potentially interested will meet the phlebotomist/interviewer. If a subject refuses to participate, then he is given the "Questions for Decliners" form and no further contact is made. The study coordinator explains the study, determines eligibility, obtains informed consent, and if appropriate administers a questionnaire, withdraws 45 cc of blood, collects buccal cells, obtains nail clipping and a urine sample in collaboration with the GCRC laboratory. As the patients await their examination in the clinic, they are accompanied by the phlebotomist/interviewer who helps them with orientation in the building etc. This gives also opportunity to answer the preliminary questions and to set a time for the full questionnaire/sample collection. This method worked well in our previous studies.

Controls are obtained from visitors accompanying other patients to the hospital. The interviewer identifies potential candidates, investigates their willingness to participate, and screens for eligibility using a script (Script 2-Control Recruitment in Clinic Area) and the eligibility screening form. The subjects usually accompany a person to the hospital on a regular basis. These controls are easily contacted and typically motivated to participate. The interviewer creates a list of willing, eligible controls and recruits from the list to the study when a match is identified. The controls are unbiased with respect to geography and socioeconomic status because they come to the hospital from the same geographic referral area as the cancer cases. In addition, controls can be obtained from neighbors and friends of the patients. Each patient can nominate up to 5 people living in the same geographical area and of the same race and age (within 5 years). The patients are asked to verify with the nominees about their agreement to be contacted by the phlebotomist/interviewer. A random drawing from the list of candidates will be performed and a candidate will be contacted. Up to three phone calls will be placed. If the subject does not return the phone calls, then it is assumed that he is uninterested in participating. In the event that a subject cannot be reached by phone, he will be contacted by mail. In case of refusal, next candidate is then randomly selected from the list of nominees. An attempt is made to collect information on age, race, smoking and drinking history of those who refuse to participate to determine whether they differ from participants demographically or by exposures. If a matching control cannot be found among the nominees, a match is identified from the pool of all eligible controls in the study. The phlebotomist/interviewer works from a list of the cases that have been enrolled up to that time, so that he/she can identify appropriate matches. Eligibility of interested controls to participate is determined over the phone by the phlebotomist/interviewer according to the telephone script. The interested candidates are invited to the Georgetown Hospital to finish a full questionnaire, donate a 45cc blood sample, a sample of buccal cells, and a sample of urine. PSA will be tested by the GCRC for all controls to exclude misclassification. Controls with PSA > 2.5ng/ml will be referred to a clinician for a follow-up testing. In this way, we obtain controls individually matched on race and age (within 5 years). Informed consent is obtained at the time of interview.

It should be noted that representatives of the U.S. Army Medical Research and Materiel Command are eligible to review research records as part of their responsibility to protect human subjects in research. Also, if any changes to the protocol or consent form are made, they are to be reviewed and approved by the Human Subjects Research Review Board prior to implementation.

Reporting of Serious and Unexpected Adverse Events:
Unanticipated problems involving risk to subjects or others, serious adverse events related to participation in the study, and all study-related subject deaths will be promptly reported by phone (301-619-2165), by email (hsrrb@det.amedd.army.mil), or by facsimile (301-619-7803) to the Army Surgeon General's Human Subjects Research Review Board (HSRRB). A complete written report will follow the initial telephone call. In addition to the methods above, the complete report can be sent to the U.S. Army Medical Research and Materiel Command, ATTN:MCMR-ZB-QH, 504 Scott Street, Fort Detrick, Maryland 21702-5012."

Procedures: Subjects are identified by review of appointment logs and discussion with doctors. Subjects are contacted during their visit to the clinic (patients), in the clinic waiting areas (controls), or by phone (controls nominated by the patient). The phlebotomist/interviewer assists the patient during his visit to the hospital, determines eligibility, explains the study and obtains informed consent, administers the questionnaire and collects 45cc of blood, buccal cells, nail clipping and a sample of urine together with the GCRC laboratory. The interviewers are trained through the GCRC in how to administer and properly complete the questionnaire. Dietary exposures (high fat etc.) will be assessed using the well-validated questionnaire developed by Dr. Gladys Block, NCI, NIH. Phlebotomy is performed by trained phlebotomists. There will be a single blood draw, using these tubes in the following order: two 7 ml green top tubes, two 7 ml plain red top tubes, one 10 ml yellow top tubes, and one 7 ml purple top tube. Only a portion of the collected samples is used for the currently planned specific aims. The remainder of the samples is aliquotted and frozen at -70^{o}C for future studies. There will be blood for multiple aliquots of buffy coat, mononuclear cells, PMNs, serum, plasma, red blood cells and clots. This strategy will allow us to test new hypotheses and assess new genetic predispositions as they are deemed worthy of study. If the subject is going to surgery, residual normal and tumor prostate tissue is placed into aliquots and snap frozen. Two samples of the normal and tumor tissues is saved, one without preservative and one with RNA later for preserving RNA. Tumor tissue is also fixed in formalin and ethanol. When available from surgery, normal cells are collected to establish primary cell cultures. If a subject is not going to surgery, but the subject had surgery at the University, then tumor blocks are requested from the LCC histopathology core. Medical records are reviewed to obtain pathological and clinical data. If a subject chooses to withdraw from the study, the link between his identity and the research study will be destroyed. Also, his biological samples will be discarded. However, the results of any finished analysis and or published result will be kept to preserve the integrity of the study.

Laboratory Methods: All the methods follow an established protocol. The mutagen sensitivity, comet assay, and apoptosis are carried out on short-term (3 day) cultured human lymphocytes exposed to bleomycin (2). The samples of isolated DNA for dGMDA quantification are sent to outside collaborators for analysis. These samples will contain only the identifier code so that there is no possibility to disclose personal information. The dGMDA is quantified by gas chromatography/negative chemical ionization mass spectrometry (25). Genetic polymorphisms are analyzed by PCR-RFLP as described (12)(19). Mutational spectra of p53 are analyzed in isolated DNA by the affymetrix chip in the laboratory of Dr. Shields (26).

Statistical Power: The present proposal intends to study 300 prostate cancer cases and 300 matched controls. The matched-pairs design increases statistical power to detect a meaningful relative risk since matched-pairs data would gain relative efficiency in estimation. Suppose the hypothesis of interest is that having a certain biomarker (e.g. mutagen sensitivity) increases the probability of developing prostate cancer, with the null hypothesis being that such probability is the same with or without the biomarker. Let p be the population frequency of having such biomarker, and let r be the relative risk defined as the ratio of the frequency of prostate cancer with the biomarker to the frequency of prostate cancer without the biomarker. Then for r=2.5, the statistical power with 5% level of significance (two-sided) will be 84%, 89%, and 93%, respectively, if p=20%, 25%, and 30%, accordingly. In our case, for example, the

frequency of mutagen sensitive subjects in the population was estimated as 20% (6) and the *XRCC1* 'at risk' allele as 25% in the general population (19). The statistical power would be relatively lower when the comparison is controlled by other factors such as race. It should be noted that tests of effect modification or associations are exploratory, and the study was not designed to have optimal power for those analyses. All the analyses will be performed using the Statistical Analysis System (SAS) and S-plus statistical software packages.

References:

1. Berwick, M. and Vineis, P. Markers of DNA Repair and Susceptibility to Cancer in Humans: an Epidemiologic Review. J Natl Cancer Inst. 6-7-2000;92(11):874-97.
2. Wu, X., Gu, J., Patt, Y., Hassan, M., Spitz, M. R., Beasley, R. P., and Hwang, L. Y. Mutagen Sensitivity As a Susceptibility Marker for Hepatocellular Carcinoma. Cancer Epi.Biom.Prev. 1998;7(7):567-70.
3. Wei, Q., Spitz, M. R., Gu, J., Cheng, L., Xu, X., Strom, S. S., Kripke, M. L., and Hsu, T. C. DNA Repair Capacity Correlates With Mutagen Sensitivity in Lymphoblastoid Cell Lines. Cancer Epidemiol Biomarkers Prev. 1996;5(3):199-204.
4. Cloos, J., Nieuwenhuis, E. J., Boomsma, D. I., Kuik, D. J., van der Sterre, M. L., Arwert, F., Snow, G. B., and Braakhuis, B. J. Inherited Susceptibility to Bleomycin-Induced Chromatid Breaks in Cultured Peripheral Blood Lymphocytes . J.Natl.Cancer Inst. 7-7-1999;91(13):1125-30.
5. Scherr, D. S., Vaughan, E. D., Wei, J., Chung, M., Felsen, D., Allbright, R., and Knudsen, B. S. BCL-2 and P53 Expression in Clinically Localized Prostate Cancer Predicts Response to External Beam Radiotherapy. J Urol. 1999;162(1):12-6.
6. Lipponen, P. and Vesalainen, S. Expression of the Apoptosis Suppressing Protein Bcl-2 in Prostatic Adenocarcinoma Is Related to Tumor Malignancy. Prostate 6-15-1997;32(1):9-15.
7. Stapleton, A. M., Zbell, P., Kattan, M. W., Yang, G., Wheeler, T. M., Scardino, P. T., and Thompson, T. C. Assessment of the Biologic Markers P53, Ki-67, and Apoptotic Index As Predictive Indicators of Prostate Carcinoma Recurrence After Surgery. Cancer 1-1-1998;82(1):168-75.
8. Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. Human DNA Repair Genes. Science 2001;291:1284-1289.
9. Shen, M. R., Jones, I. M., and Mohrenweiser, H. Nonconservative Amino Acid Substitutions Exist at Polymorphic Frequency in DNA Repair Genes in Healthy Humans. Can. Res. 2-15-1998;58:604-8.
10. Boiteux, S.,and Radicella, J. P. The Human OGG1 Gene: Structure, Function, and its Implication in the Process of Carcinogenesis. Arch. Biochem. Biophys. 2000;377(1):1-8.
11. Kohno, T., Shinmura, K., Tosaka, M., Tani, M., Kim, S. R., Sugimura, H., Nohmi, T., Kasai, H., and Yokota, J. Genetic Polymorphisms and Alternative Splicing of the hOGG1 Gene that is Involved in Repair of 8-hydroxyguanine in Damaged DNA. Oncogene. 1998;16(25):3219-3225.
12. Sugimura, H., Kohno, T., Wakai, K., Nagura, K., Genka, K., Igarashi, H., Morris, B. J., and Yokota J. hOGG1 Ser326Cys Polymorphism and Lung Cancer Susceptibility. Cancer Epidemiol Biomarkers Prev. 1999;8(8):669-674.
13. Wikman, H., Risch, A., Klimek, F., Schmezer, P., Spiegelhalder, B., Dienemann, H., Kayser, K., Schulz, V., Drings, P., and Bartsch, H. hOGG1 Polymorphism and Loss of Heterozygosity (LOH): Significance for Lung Cancer Susceptibility in a Caucasian Population. Int J Cancer. 2000;88(6):932-937.
14. Duell, E. J., Wiencke, J. K., Cheng, T. J., Varkonyi, A., Zuo, Z. F., Ashok, T. D., Mark, E. J., Wain, J. C., Christiani, D. C., and Kelsey, K. T. Polymorphisms in the DNA Repair Genes XRCC1 and ERCC2 and Biomarkers of DNA Damage in Human Blood Mononuclear Cells [Published Erratum Appears in Carcinogenesis 2000 Jul;21(7):1457]. Carcinogenesis 2000;21(5):965-71.
15. Sturgis, E. M., Castillo, E. J., Li, L., Zheng, R., Eicher, S. A., Clayman, G. L., Strom, S. S., Spitz, M. R., and Wei, Q. Polymorphisms of DNA Repair Gene XRCC1 in Squamous Cell Carcinoma of the Head and Neck. Carcinogenesis 1999;20(11):2125-9.
16. Abdel-Rahman, S. Z., Soliman, A. S., Bondy, M. L., Omar, S., El Badawy, S. A., Khaled, H. M., Seifeldin, I. A., and Levin, B. Inheritance of the 194Trp and the 399Gln Variants of the DNA Repair Gene XRCC1 Are Associated With Colorectal Carcinoma in Egypt. Cancer Lett. 2000;159(1):79-86.
17. Divine, K. K., Gilliland, F. D., Crowell, R. E., Stidley, C. A., Bocklage, T. J., Cook, D. L., and Belinsky,

S. A. The XRCC1 399 Glutamine Allele Is a Risk Factor for Adenocarcinoma of the Lung. Mutat.Res 1-5-2001;461(4):273-8.

18. Block, G., Patterson, B., and Subar, A. Fruit, Vegetables, and Cancer Prevention: a Review of the Epidemiological Evidence. Nutr.Cancer 1992;18:1-29.
19. Lunn, R. M., Langlois, R. G., Hsieh, L. L., Thompson, C. L., and Bell, D. A. XRCC1 Polymorphisms: Effects on Aflatoxin B1-DNA Adducts. Cancer Res. 6-1-1999;59(11):2557-61.
20. Dogliotti, E., Hainaut, P., Hernandez, T., D'Errico, M., and DeMarini, D. M. Mutation spectra resulting from carcinogenic exposure: from models to cancer genes. Rec. Res. Can. Res., *154*: 97-124, 1998.
21. Levine, A. J. p53, the cellular gatekeeper for growth and division. Cell, *88*: 323-331, 1997.
22. DeBenedetti VMG, Travis LB, et al. p53 mutations in lung cancer following radiation therapy for hodgkin's disease. Cancer Epidemiol Biomark Prev 1996; 5:93-98.
23. Vahakangas KH, Samet JM, et al. Mutations of p53 and ras genes in radon-associated lung cancer from uranium miners. Lancet, 1992; 339:576-580.
24. Chi SG, deVereWhite RW, et al. p53 in prostate cancer: frequent expressed transition mutations. JNCI 1994; 86:926-933.
25. Otteneder M, Plastaras JP, and Marnett LJ. Reaction of malondialdehyde DNA adducts with hydrazines-a facile assay for quantification of malondialdehyde in DNA. Chem Res Tox. 2002 Mar; 15:312-318
26. Wolf P, Hu YC, Doffek K, Sidransky D, Ahrendt SA. O(6)-Methylguanine-DNA methyltransferase promoter hypermethylation shifts the p53 mutational spectrum in non-small cell lung cancer. Cancer Res. 2001 Nov 15;61(22):8113-7.

8. Indicate what you consider to be the risks to subjects and indicate the precautions to be taken to minimize or eliminate these risks. Justify the need for a placebo control group if one is included in this study. Where appropriate, describe the data monitoring procedures that will be employed to ensure the safety of subjects. Use additional sheets as needed.

There are minimal risks for this study. The only invasive procedure is phlebotomy. This may cause a bruise on the arm from the needle stick and possibly an infection. These risks are minimized through proper techniques for phlebotomy and the trained staff is experienced in reducing discomfort to patients. The actual surgery or clinical practices related to the prostate cancer will not be altered for this study.

Section Four: Selection of Subjects and the Informed Consent Process

9. Indicate whether this project involves any of the following subject populations?
   - ❑ Children (Children are defined by local law as anyone under age 18.)
   - ❑ Prisoners
   - ❑ Pregnant women
   - ❑ Cognitively impaired or mentally disabled subjects
   - ❑ Economically or educationally disadvantaged subjects

If you indicated any of the above, in the space below, please describe what additional safeguards will be in place to protect these populations from coercion or undue influence to participate. (Use additional sheets as needed.)

10. Describe how subjects will be recruited and how informed consent will be sought from subjects or from the subjects' legally authorized representative. If children are subjects, discuss whether their assent will be sought and how the permission of their parents will be obtained. Use additional sheets as needed.

This is a study of prostate cancer risk factors that enrolls newly diagnosed, incident prostate cancer cases from the Departments of Medicine and Oncology, Radiation Medicine, and Urology at the Georgetown University Medical Center. The eligible patients donate their time for a questionnaire; blood and urine samples; buccal swabs; nail clipping; and unneeded normal and tumor prostate tissue. Subjects are eligible and will be enrolled even if they are not having a surgery or biopsy and if no tissues are available. Subjects older than 18 years of age at all stages of presentation are included. No subject is excluded based on minority status. Subjects with psychiatric disorder or any other reason that precludes understanding the informed consent are excluded for ethical reasons. The phlebotomist/interviewer conducts a brief initial 15 minute interview in order to explain the study, determine eligibility, and explain the informed consent. If a subject refuses to participate, then no further contact is made. If appropriate, the phlebotomist/interviewer administers a structured forty five minute interview that establishes demographic characteristics, family history of cancer, dietary habits, tobacco and alcohol use, occupational exposures, and history of vasectomy. This interview can be done at any time up to two months after initiation. The phlebotomist/interviewer will also withdraw 45 cc of blood, collect buccal cells, obtain nail clipping and a urine sample in collaboration with the GCRC laboratory at Georgetown University.

Controls are obtained from visitors accompanying other patients to the hospital. The interviewer identifies potential candidates, investigates their willingness to participate, and screens for eligibility using a one-page form. The interviewer creates a list of willing, eligible controls and recruits from the list to the study when a match is identified. In addition, controls can be obtained from neighbors and friends of the patients. Each patient can nominate up to 5 people living in the same geographical area and of the same race and age (within 5 years). The patients are asked to verify with the nominees about their agreement to be contacted by the phlebotomist/interviewer. The controls are randomly selected from the list of candidates and contacted by the interviewer. Up to three phone calls are placed. If the subject does not return the phone calls, then it is assumed that he/she is uninterested in participating. In case of refusal, next candidate is randomly selected from the list of nominees. An attempt is made to collect information on age, race, smoking and drinking history of those who refuse to participate to determine whether they differ from participants demographically or by exposures. A subsequent meeting with the matching

control is scheduled. During this meeting, the interviewer explains the study in detail and obtains informed consent. A full length questionnaire as well as blood , buccal, urine, and nail-clipping samples are obtained. The samples or questionnaire can be obtained also at a later visit up to two month following the initial contact if this is more convenient for the participant.

11. Will subjects receive any compensation for participation in cash or in kind?
    - √ Yes. If so, please describe amount or kind of compensation in the space below.
    - ❑ No.

Patients will not be compensated. Controls will receive free PSA test if needed and $25 for parking if study funds permit.

Section Five: Privacy and Confidentiality of Data and Records

12. Will identifiable, private, or sensitive information be obtained about target the subjects or other living individuals? Whether or not such information is obtained, describe the provisions to protect the privacy of subjects and to maintain the confidentiality of data. Use additional sheets as needed.

There are minimal risks of disclosure of sensitive information in this study, but there is always the risk that genetic or other risk factor data might be obtained by the subject or a third party. However, it is important to realize that the genes studied herein are low penetrant. We study only common genetic polymorphisms in DNA repair genes and somatic mutations in p53; we do not study familial germ line mutations. This risk of disclosure will be minimized by the confidentiality and protection of privacy procedures described below.

Protection of privacy of participants in genetic studies is of the utmost importance. Study subject's confidentiality is maintained at all times. Subjects are assigned unique study numbers. These unique study numbers are linked to the subject's identifier information in a database and on the hard copy of the Identifier Sheet. This information is secured by Dr. Goldman in his office separate from the laboratory. The database requires at least two levels of security (i.e. passwords), which allows only authorized individuals to access the information. The Identifier Sheets are physically separated from the questionnaire and stored in a locked cabinet. The questionnaire retains only the unique study number. Biological samples are labeled with the unique study number and no other identifier information. No identifier information that can be linked to study results or other data will leave Dr. Goldman's premises.

Identifier information for non-participants (refusers and ineligibles) is recorded in order to avoid recontact. This information is stored in a database with at least two levels of security (i.e. passwords), which allows only authorized individuals to access the information. A log will automatically note who accesses the information and what was accessed. Unique study number for non-participants is also assigned; this is used for tracking reasons. Two databeses are maintained. The first includes the Contact Database and includes identifier information. It will record if subjects refused, were ineligible, or are participants. If participants, it will record when the interview occurred or will occur, the outcome, and track sample handling. For refusers and ineligibles, it will record that their data was entered into the Refusal and Ineligible database. The Refusal and Ineligible database will record data and why the subject was ineligible. This database does not contain identifier information.

I certify that the information furnished concerning the procedures to be taken for the protection of human subjects is correct. I will seek and obtain prior approval for any modification in the protocol or informed consent document and will report promptly any unexpected or otherwise significant adverse effects encountered in the course of this study.

I certify that all individuals named as consultants or co-investigators have agreed to participate in this study.

_____ Printed/Typed Name of Investigator	_____ Telephone number
_____ Signature of Investigator	_____ Date
Department Chair: ❑ Approved ❑ Disapproved _____ Printed/Typed Name _____ Signature of Department Chair	_____ Telephone Number _____ Date

If more than one department or administrative unit is participating in the research and/or if the facilities or support of another unit, e.g., nursing, pharmacy, or radiation therapy, are needed, then the chair or administrative official of each unit must also sign this application.

_____ Authorized Signature and Title	_____ Date
_____ Authorized Signature and Title	_____ Date
_____ Authorized Signature and Title	_____ Date
_____ Authorized Signature and Title	_____ Date
_____ Authorized Signature and Title	_____ Date

Section Six: Attachments

Please attach the following items in order for the IRB to review your research.

1. 24 copies of this IRB Application form
2. The informed consent document (24 copies)
3. Any recruitment notices or advertisements (24 copies)
4. Any research survey instruments, psychological tests, interview forms, or scripts to be used (24 copies).
5. Certificate of Completion of Education in the Protection of Human Research Subjects
6. Investigator's qualifications (CV, biosketch, or Form 1572, if available)
7. Investigator's Brochure from the sponsor, if applicable (5 Copies)
8. Research protocol and sample consent document from the sponsor or Cooperative Group, if applicable (5 copies)
9. Grant application, if applicable (2 copies)

Investigator's Brochure (where applicable)

The Investigator's Brochure must contain the following information. If it does not contain the information, then please attach a separate sheet of paper to address the item.

(a) Name of drug under study.
(b) Source of the drug.
(c) Experience with the drug in humans, including doses tested, toxicity observed, minimal toxic dose, pharmacokinetic data (absorption, elimination, metabolism, etc.).
(d) Description of toxicity in humans.
(e) Procedures for minimizing adverse reactions and dealing with those that might occur.

MEDICAL RECORDS RELEASE AND
GENERAL AUTHORIZATION TO USE AND DISCLOSE HEALTH INFORMATION FOR RESEARCH

I agree to allow Dr. Goldman and his staff (together called "Researchers"), as well as the study sponsor, Lombardi Cancer Center of Georgetown University, others working with the sponsor to do the research (together called "Sponsor"), and the other people or companies listed below, to use and give my personal health information that identifies me for the reason described in the Informed Consent Form used for this study and as needed to conduct the research. I also agree to allow Georgetown University Hospital, my doctors and my other health care providers, and others who generate or use my health information, to give my health information in my medical or other records to the Researchers and Sponsor for the purposes described below and in the Informed Consent Form used in this study. [IRB Project # 03013 and Project Full Title: The Molecular Epidemiology of Prostate Cancer]

1. The health information that may be used for this study includes:

☒ All my personal information made or collected during the research described in the Informed Consent Form for this study; and

☒ All my personal information in my medical records requested by the Researchers to be able to do the research described in the Informed Consent Form for this study.

OR

☐ The following information: _____

_____

2. The person(s), class(es) of persons, and/or organizations (companies) who may use, give and receive the above information include*:

☒ Every research site for this study, including the hospital, and including each site's research staff, medical staff and administrative staff;

☒ Health care providers who provide services to me in connection with this study;

☒ Laboratories and other individuals and organizations that look at my health information in connection with this study, in agreement with the study's protocol;

☐ The Sponsor and the people and companies that they use to watch over how the study is managed, run, or do the research as described above;

☐ The United States Food and Drug Administration (FDA) and other Federal or State Agencies that watch over the safety of the study and how the study is managed or run;

☒ The members and staff of the Institutional Review Board(s) or Ethics Committee(s) that approves this study;

☒ The Principal Investigator, other Investigators, Study Coordinators, and all administrative staff in charge for doing all the work for the study and other research activities;

☐ The Patient Advocate or Research Ombudsman (people who watch out for my best interest):

_____

☐ Data Safety Monitoring Boards (a group of people who examine the medical information during the study) and other government agencies or review boards who watch over the safety, success and how the research is done.

☐ Others:_____

_____

**If, during the course of the research, one or more of the companies or institutions above merges (becomes one company) or is bought by another company, this Authorization will remain valid.*

3. Once my health information has been given to one of the person(s), class(es) of persons, and/or organizations (companies) listed above, there is the possibility that federal privacy laws (laws that protect the privacy to my personal health information) may no longer protect it from being given to

another person, class of persons, and/or company. However, the Researchers and Sponsor [may agree/have agreed] to further protect my health information by using and disclosing it only for the research purposes described in the Informed Consent Form and as allowed by me in this Authorization (agreement). Also, the Researchers and Sponsor [may agree/have agreed] that no publication or presentation of the research will reveal my identity without my separate specific written permission and authorization (agreement). These limitations, if agreed to by the Researcher and Sponsor, continue even if I revoke (take back) this Authorization (agreement).

4. **Once information that could be used to identify me has been removed and my information is no longer identifiable (connected to my identity) under federal regulations,** the information that remains is no longer protected by this Authorization (agreement) and may be used and given by the Researchers and Sponsor as permitted by law to others, including for other research reasons.

5. **I understand that:**
   - I have the right to refuse to sign this Authorization (agreement). While my health care outside the study, the payment for my health care, and my health care benefits will not be affected if I do not sign this form, I will not be able to participate in the research described in this Authorization (agreement) and will not receive treatment as a study participant if I do not sign this form.
   - I may change my mind and revoke (take back) this Authorization (agreement) at any time. To take back this Authorization (agreement), I must write to: Allison Pollock, Lombardi Cancer Center, Lower Level Room S-180, Georgetown University, Box 571472, Washington, DC 20057-1472 . However, if I take back this Authorization (agreement), I may no longer be allowed to participate in the research. Also, even if I take back this Authorization (agreement), the information already obtained may remain a part of the research as necessary to preserve the integrity of the research study.

6. **This Authorization (agreement) does not have an expiration (ending) date.**

7. **I will be given a copy of this Authorization (agreement) after I have signed it.**

8. **I acknowledge that I have received or declined the pamphlet with the MedStar Health Notice of Privacy Practices and that this form is available for me to take with me.**

_____ _____

Signature of participant or participant's legal representative | Date

_____ _____

Printed name of participant or participant's representative | Representative's authority to sign for participant

For Internal Use Only

Signature/acknowledgement of receipt of Notice of Privacy Practices not obtained because:

☐ Emergency

☐ Patient/Patient Representative declined to sign

☐ Patient/Patient Representative unable to sign

_____

MRI Representative

Molecular Epidemiology of Prostate Cancer (Case/Control)

Principal Investigator: Radoslav Goldman, Ph.D.
Department of Oncology
Lombardi Comprehensive Cancer Center
Georgetown University Medical Center
LCC, LL (S) Level, S183
3800 Reservoir Road, NW
Washington, DC 20057
Tel: (202) 687 9868
Fax: (202) 687 1988
email: rg26@georgetown.edu

TABLE OF CONTENTS

Date of Interview ____ ____ _____ MM DD YYYY	Time of Interview _____ ☐ 1 AM _____ ☐ 2 PM

Interviewer	Interviewer Signature

Study ID/ Site ID	LCC Number

MRN	Control? Yes_____ No_____

Reviewers initials	Date reviewed ____ ____ _____ MM DD YYYY
Coders initials	Dated coded ____ ____ _____ MM DD YYYY
First Entry initials	Date entered ____ ____ _____ MM DD YYYY
Second entry initials	Date entered ____ ____ _____ MM DD YYYY

Date Samples Collected _____	ID label
Blood ☐ ___ yellow ___ red ___ green ____purple	
Mouthwash ☐ ___	
Urine ☐ ___	
Toenail ☐ ___	
Tissue ☐	
PSA ☐	

Your answers to the following questions are very important to us. Please answer them as truthfully as possible. Also, please remember that you do not have to answer any question that makes you feel uncomfortable.

A. IDENTIFIER SHEET

A1. What is your name? _____ / _____ / _____
First Middle Last

A2. Could your medical records be under a different name? If so, what name?

_____ / _____ / _____
First Middle Last

A3. What is your date of birth? ___ ___ / ___ ___ / ___ ___ ___ ___
MM DD YYYY

A4. What is your address?

_____
Street Apt. No.

_____ ____ ____ __ __ __ __ __ - __ __ __ __
City State Zip Code

_____
Country

A5. What is your telephone number? Home: (__ __ __) __ __ __ - __ __ __ __

Work:(__ __ __) __ __ __ - __ __ __ __ Ext. __ __ __ __

Email _____

IDENTIFIER SHEET ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

B. DEMOGRAPHIC INFORMATION

Now I would like to ask you some general information about yourself.

B1. What is your marital status?

( )1 Widowed
( )2 Married or living as married
( )3 Divorced
( )4 Separated
( )5 Single, never married

B2. Which of these categories best describes you?

( )1 White
( )2 Black or African American
( )3 Asian
( )4 Native Hawaiian or Other Pacific Islander
( )5 Other Specify_____

B3. What country or continent were you born in?

( )1 United States	( )2 Africa	( )3 Europe
( )4 Caribbean/West Indies	( )5 Asia	( )6 South America
( )7 Middle East	( )8 Canada	( )9 Australia
( )10 United Kingdom	( )11 Central America	( )12 Other_____

B4. If you moved from here, at what age did you move?_____

B5. What was the highest level of education you completed (don't read choices).

( )1 Less than 8th grade ( )2 Less than high school ( )3 High school graduate
( )4 Less than 4 years of college ( )5 College (4 years completed)
( )6 Graduate/professional coursework or degree

B6. In what religion were you raised?

( )1 Protestant	( )2 Catholic	( )3 Muslim
( )4 Jewish	( )5 None	( )6 Other Specify _____

If Jewish, are you Ashkenazi? _____yes _____no

B7. What is your current level of household income per year (read choices)?

( )1 Less than $25,000
( )2 $25,001 - $50,000
( )3 $50,001 - $100,000
( )4 $100,001- $150,000
( )5 Greater that $150,000
( )8 Don't know

B8. How many people are currently supported in your household? ____ ____

DEMOGRAPHIC INFO ( )1 Very Good ( )2 Good ( )3 Fair ( )4 Poor

C. MEDICATIONS

C1. Now I have some questions about any prescription medication you may have taken.

Drugs	C1.Have you ever taken (DRUG)?	C2. In what year did you first take (DRUG)?	C3. For how long did you take (DRUG)?	C4. How often did you take (DRUG) per day or per week?
a. Propecia used to treat baldness?	YES..... 1 → NO..... 2 (b)	|___|___|___|___|	|___|___| MONTHS..... 1 YEARS..... 2	|___|___| PER DAY..... 1 PER WEEK..... 2
b. Proscar or fenasteride used to treat prostate disease?	YES..... 1 → NO..... 2 (c)	|___|___|___|___|	|___|___| MONTHS..... 1 YEARS..... 2	|___|___| PER DAY..... 1 PER WEEK..... 2
c. Luprone or Zolodex used to treat prostate disease?	YES..... 1 → NO..... 2 (d)	|___|___|___|___|	|___|___| MONTHS..... 1 YEARS..... 2	|___|___| PER DAY..... 1 PER WEEK..... 2
d. Flutamide also called Eulexin; or Nilandron; or Casodex used to treat prostate disease?	YES..... 1 → NO..... 2 (e)	|___|___|___|___|	|___|___| MONTHS..... 1 YEARS..... 2	|___|___| PER DAY..... 1 PER WEEK..... 2
e. Urinary Obstruction Control Drugs. (Calcium Channel Blockers) (eg: Calan, Isoptin, Covera-HS, Varelen, Cardene, Adalat, Procardia, Cardura, Hytrin, Flomax,)	YES..... 1 → NO..... 2 (f)	|___|___|___|___|	|___|___| MONTHS…… 1 YEARS…… 2	|___|___| PER DAY…….. 1 PER WEEK...…… 2
f. Viagra, Cialis, Levitra. Which one? _____	YES..... 1 → NO..... 2 (C5)	|___|___|___|___|	|___|___| MONTHS…… 1 YEARS……... 2	|___|___| PER DAY…….. 1 PER WEEK…… 2 OCCASIONALLY… 3

C5. Now I have some questions about supplements and other drugs some men take.

OTHER DRUGS AND SUPPLEMENTS	C5. Did you ever take (SUPPLEMENT)?	C6. In what year did you start to take (SUPPLEMENT)?	C7. How long did you take (SUPPLEMENT)?	C8. How often did you take (SUPPLEMENT) per day or per week?
a. DES (Diethyl stilbesterol)	YES……..1 → NO.....2 (b)	|___|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
b. Prostate Healthcare Drugs (ex: PC SPES, Saw Palmetto, Dayto, Homemix, Yohimbe, Damiana leaf) Which one? _____	YES.....1 → NO.....2 (c)	|___|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2

c. Lasix	YES.....1→ NO.....2(d)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
d. Lycopene	YES.....1 → NO.....2 (e)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
e. Selenium	YES.....1 → NO.....2 (f)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
f. Vitamin E	YES.....1 → NO.....2(g)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
g. Body Building or performance enhancing steroids.(DHEA, 19-Nor/androstenedione) Which one?_____	YES.....1 → NO.....2(h)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
h. Statins or Cholersterol lowering drugs (ex. Lipitor, Zocor, Mevacor) Which one?_____	YES.....1 → NO.....2 (i)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
i. Cox-2 Inhibitors (Celebrex, Vioxx, Bextra)	YES.....1 → NO.....2 (j)	|___|___|___|___|	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
j.Multivitamin. Which one(s)? _____ _____	YES.....1 → NO.....2 (C9)	|___|___|___|___	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
k. Other Vitamins. Which one(s)? _____	YES.....1 → NO.....2 (C9)	|___|___|___|___	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
_____	YES.....1 → NO.....2 (C9)	|___|___|___|___	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2
_____	YES.....1 → NO.....2 (C9)	|___|___|___|___	|__|__| MONTHS….1 YEARS…….2	|__|__| PER DAY 1 PER WEEK 2

C9. Have you ever taken non-steroidal anti-inflammatory drugs (NSAIDs) such as Aspirin, Bufferin, Excedrin, Advil, Motrin, Nasproxsyn, and Ibuprofen (Tylenol is not an NSAID)?

( )$_0$ No (Skip to C12) ( )$_1$ Occasionally (Skip to C12) ( )$_2$ Weekly (Skip to C12) ( )$_3$ Daily

*Which one?*_____

C10. For what reason did you take NSAIDs?

( )$_0$ Headache ( )$_1$ Heart disease ( )$_2$ Stroke

( )$_3$ Arthritis ( )$_4$ Other_____(please specify)

C11. If you have taken NSAIDs **on a daily basis**, I would like to ask you about these periods during different times of your life. (Fill in table below)

Action	Period 1	Period 2	Period 3	Period 4	Period 5
a. In what year did you start taking these drugs?					
b. How many or how much did you take per day?	_____ ( )pills ( )mg	_____ ( )pills ( )mg	_____ ( )pills ( )mg	_____ ( )pills ( )mg	_____ ( )pills ( )mg
c.Which type or brand did you use?	_____	_____	_____	_____	_____
d. Did you continue to take this, stop or Δ your pattern for more than 6 months?	$(\ )_0$ continued $(\ )_1$ stopped $(\ )_2$ pattern Δ	$(\ )_0$ continued $(\ )_1$ stopped $(\ )_2$ pattern Δ	$(\ )_0$ continued $(\ )_1$ stopped $(\ )_2$ pattern Δ	$(\ )_0$ continued $(\ )_1$ stopped $(\ )_2$ pattern Δ	$(\ )_0$ continued $(\ )_1$ stopped $(\ )_2$ pattern Δ
e. Year you stopped taking NSAIDS or Δ your pattern for >6 months?	_____ If this is a Δ of pattern, ⇒C2a	_____ If this is a Δ of pattern, ⇒C3a	_____ If this is a Δ of pattern, ⇒C4a	_____ If this is a Δ of pattern, ⇒C5a	_____
f. Did you start NSAIDS again?	$(\ )_0$ no ⇒C6 $(\ )_1$ yes ⇒C2a	$(\ )_0$ no ⇒C6 $(\ )_1$ yes ⇒C2a	$(\ )_0$ no ⇒C6 $(\ )_1$ yes ⇒C2a	$(\ )_0$ no ⇒C6 $(\ )_1$ yes ⇒C2a	$(\ )_0$ no $(\ )_1$ yes

C12. Have you taken any other prescription or non-prescription medications within the last year?

$(\ )_0$ No **(Skip to D)** $(\ )_1$ Yes

C13. Which ones?

Name of Medication	Date began?	Date finished?	Reason for taking?	Notes

MEDICATIONS $(\ )_1$ Very Good $(\ )_2$ Good $(\ )_3$ Fair $(\ )_4$ Poor

D. SMOKING HISTORY

Now I have some questions about smoking.

D1. Have you ever smoked a total of 100 cigarettes or more in your lifetime?
( )$_0$ No **(Skip to E1)** ( )$_1$ Yes

D2. Did you ever smoke cigarettes regularly, at least one cigarette per day for six months or longer?
( )$_0$ No **(Skip to E1)** ( )$_1$ Yes

D3. How old were you when you first started smoking regularly? |___|___|
AGE STARTED

D4. Do you smoke cigarettes regularly now?
( )$_0$ No ( )$_1$ Yes (**Skip to D6)**

D5. How old were you when you stopped smoking regularly? |___|___|
AGE STOPPED

D6. In total, how many years have you smoked or did you smoke regularly (please subtract out years you did not smoke)?
|___|___|
YEARS

D7. Thinking about all the years when you smoked regularly, how many cigarettes did you usually smoke in a day?
|___|___|___|
CIGARETTES/DAY

D8. During your childhood, until you were 18, did anyone in your home smoke? (do not include this if smoking was done only outside the home). ( )$_0$ No (skip to D10) ( )$_1$ Yes

D9. How many people smoked in your home during your childhood? _____

D10. As an adult, does/did your spouse or partner or anyone else smoke in your home? (do not include this if smoking is/was done only outside the home). ( )$_0$ No ( )$_1$ Yes

D11. How many people smoked in your home during your adulthood? _____

D12. Do/Did you work in a place where co-workers smoked in your immediate area? ( )$_0$ No ( )$_1$ Yes

D13. For how many years were you working at a job where people smoked regularly in your immediate work area _____

SMOKING HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

E. ALCOHOL HISTORY

E1. Did you ever drink any alcoholic beverages, such as beer, wine or hard liquor, on a regular basis, that is, at least once a week for 6 months or longer?

( )$_0$ No **(Skip to F1)** ( )$_1$ Yes

E2. How old were you when you started drinking regularly? |___|___|
AGE STARTED

E3. Do you still drink regularly now? ( )$_0$ No ( )$_1$ Yes **(Skip to E5)**

E4. How old were you when you stopped drinking regularly? |___|___|
AGE STOPPED

E5. In total, for how many years have you or did you drink regularly? Please subtract out the years when you didn't drink regularly. |___|___|
YEARS

E6. On the average, after age 25, how many (ALCOHOLIC BEVERAGE) did you drink per week? DRINKS	E7. How many years did you drink (ALCOHOLIC BEVERAGE) regularly? YEARS
1. .....Cans or Bottles of Beer |___|___|___|	|___|___|
2. .....Glasses of Wine |___|___|___|	|___|___|
4. ..... Shots of hard liquor |___|___|___|	|___|___|

ALCOHOL HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

F. OCCUPATIONAL HISTORY

We would like some information about the types of jobs you had **for the longest period of time.**

F1. What was the complete title of this job?_____

F2. What year did you begin this job and what year did you stop? ___/___ - ___/___
mo yr mo yr

F3. Was this position a full-time or part-time job? (Full-time is 35 hours or more per week)
( )$_0$ Full-time ( )$_1$ Part-time

F4. What type of business or industry was this; that is what did this employer make or do? Please be as specific as possible._____

F5. What are/were your usual activities in this job?
_____

OCCUPATIONAL HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

G. BODY SIZE/ ANTHROPOMETRY

G1. How tall are you? |___|___| |___|___| or |___|___|
FT INCHES CM

DON'T KNOW----------988

G2. When you were about 8-9 years old, compared to other boys your age, were you ….?

Short	1
Somewhat short	2
Average height	3
Somewhat tall or	4
Tall?	5
DON'T KNOW	8

G3. When you were about 20-25 years old, compared to other men your age, were you ….?

Short	1
Somewhat short	2
Average height	3
Somewhat tall or	4
Tall?	5
DON'T KNOW	8

At what age did you reach your adult height? _____years

G4. After age 25, what has been your usual weight? |___|___|___| or |___|___|
LBS KG

DON'T KNOW 998

G5. Have you lost weight in the last 5 years? ( )$_0$ No ( )$_1$ Yes **(Skip to G8)**

G6. How much weight did you lose? |___|___|___| (IF LT 10 LBS GO TO G8)
LBS

G7. In the past 5 years, did you lose this weight without trying? ( )$_0$ No ( )$_1$ Yes

IN G8-G9, ASK EACH AGE GROUP ENDING WITH CURRENT AGE GROUP

	Age group				In the past year (prior to diagnosis)
	In 2nd to 4th grade	20-29 yrs old	40-49 yrs old	60-69 yrs old	
G8. When you were (AGE GROUP), compared with other males in the same age group were you ...?					
.....Very thin	 1	 1	 1	 1	 1
.....Somewhat thin	 2	 2	 2	 2	 2
..... Average	 3	 3	 3	 3	 3
.....Somewhat heavy	 4	 4	 4	 4	 4
..... Very heavy	 5	 5	 5	 5	 5
.....DON'T KNOW	 8	 8	 8	 8	 8
.....NOT APPLICABLE	 0	 0	 0	 0	 0
G9. What was your average weight at/in (AGE GROUP)?	|__|__|__| LBS	|__|__|__| LBS	|__|__|__| LBS	|__|__|__| LBS	|__|__|__| LBS
.....DON'T KNOW	998	998	998	998	998

G10. As an adult, what was your highest weight? |___|___|___| or |___|___|
LBS KG

G11. At what age did you first reach this highest weight? |___|___|
AGE

G12. For how many years or months were you at this highest weight? |___|___| MONTHS 1
YEARS 2

G13. When you gain weight, where on your body do you mainly tend to add the weight?

( )$_0$ don't gain weight
( )$_1$ around the waist and stomach
( )$_2$ around the hips and thighs
( )$_3$ around the chest and shoulders
( )$_4$ equally all over
( )$_5$ other (specify) _____

G14. Interviewer will ask:
Do you know your waist circumference, or pant-size? |__|__| inches

G16. How would you describe your chest hair density? ( )$_0$ thick ( )$_1$ medium ( )$_2$ thin ( )$_3$ no hairs

G17. Have you experienced any permanent hair loss from your scalp since you were twenty years old? ( )$_0$ No ( )$_1$ Yes

G18. If yes, at what age did the hair loss begin? __ __ years

G19. Interviewer: Please indicate hair thickness ( )$_0$ thick ( )$_1$ medium ( )$_2$ thin ( )$_3$ no hairs

G20. Interviewer: Please indicate hair pattern on dome
( )$_0$ no evident loss
( )$_1$ some loss
( )$_2$ patterned baldness
( )$_3$ few hairs
( )$_4$ no hairs

Baldness Some Loss

Patterned

G21. Have you ever used any hair growth products? ( )$_0$ No ( )$_1$ Yes

G22. Are you using a wig or toupee? ( )$_0$ No ( )$_1$ Yes

BODY SIZE/ANTHROPOMETRY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

H. MEDICAL HISTORY

Now I am going to ask some questions about your health.

H1. Has a doctor ever told you that you had any of the following diseases? FOR EACH *YES* RESPONSE ASK I2. FOR EACH *NO* RESPONSE GO THE NEXT DISEASE				H2. IF YES Please tell me how old you were when the disease was (first) diagnosed.
	YES	NO		AGE
a.....Peptic ulcer	1	0	(b)	a. |___|___|
b. ..... Liver cirrhosis	1	0	(c)	b. |___|___|
c..... Other liver diseases	1	0	(d)	c. |___|___|
d. .....Hepatitis B	1	0	(e)	g. |___|___|
e.....Hepatitis C	1	0	(I3)	h. |___|___|

H3. Have you ever been told by a doctor that you have diabetes or sugar diabetes?

( )$_0$ No **(Skip to I)**
( )$_1$ Yes

H4. At what age did your doctor first tell you this? __ __ years

H5. Are you now taking insulin?

( )$_0$ No **(Skip to H.8)**
( )$_1$ Yes

H6. At what age did you begin to take insulin? __ __ years

H7. For what reason do you take insulin? _____

H8. Are you now taking pills to lower you blood sugar? These are sometimes called oral agents or oral hypoglycemic agents?

( )$_0$ No **(Skip to I)** ( )$_1$ Yes

H9. At what age did you begin to take hypoglycemic agents? __ __ years

H10. For what reason do you take hypoglycemic agents? _____

MEDICAL HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

I. PROSTATE CANCER SCREENING HISTORY/UROLOGIC HEALTH

Now I'd like to ask you some questions about your urologic health.

Screening History

I1. Do you know the approximate date of your most recent examination (PSA test, DRE) for prostate cancer?
____/____/_____ _____ Don't remember ___ Never had examination (skip to I13)

I2. Was this examination performed by: _____your physician $_0$
_____a new physician who you did not know previously $_1$
_____in a free prostate cancer screening program $_2$
_____ in this study

I3. Was the prostate exam done because you were experiencing any prostate-related symptoms (e.g., urinary control, pain)? ___yes$_1$ ____no$_0$ ____don't know$_8$

I4. Was your Digital Rectal Examination abnormal? ____yes$_1$ ____no$_0$ ____don't know$_8$

I5. Were you told that your PSA was elevated? ____yes$_1$ ____no$_0$ (skip to I8) ____don't know$_8$

I6. What was your PSA value? _____(don't know=888)

I7. Did you follow up with further testing? ____yes$_1$ ____no$_0$

I8. Before this last exam , have you ever had an abnormal exam in the past (meaning that your doctor thought there was something that needed to be checked out further)? ____yes$_1$ ____no$_0$ ____don't know$_8$_____Never had exam before this last one$_9$.

I9. [IF YES] Have you had a biopsy previously? ____yes$_1$ ____no$_0$ ____don't know$_8$

a. Biopsy type	Diagnosis	Date	Hospital	Doctor
_____	_____	___/___/_____	_____	_____
_____	_____	___/___/_____	_____	_____
_____	_____	___/___/_____	_____	_____

I10. How often do you get checked out for prostate cancer?
_____every 3-6 months$_0$
_____annually$_1$
_____every 2 years$_2$
_____less often$_3$
_____don't know$_8$

I11. Approximately how many times would you say you have been checked for prostate cancer in your lifetime?
(*This would include the PSA and/or DRE)* _____(Don't know=888)

I12. Have you ever been screened in a free, mass screening program? ____yes$_1$ ____no$_0$

Urologic Health/History

I13. During a typical night, how many times do you wake up to urinate? *(For cases, please ask about a typical night during the 12 months prior to the prostate cancer diagnosis)*

( )$_0$ never (Skip to I15)
( )$_1$ once (Skip to I15)
( )$_2$ twice
( )$_3$ three times
( )$_4$ more than three times

I14. How old were you when you first began waking to urinate **more than once a night** on a regular basis?

___ ___ years

I15. Did a doctor ever tell you that you had:	Yes/No	How old were you when you were diagnosed?
a. an enlarged prostate or benign prostatic hypertrophy	( )$_0$ No ( )$_1$ Yes ( )$_8$ Don't know	
b. an inflamed prostate or prostatitis	( )$_0$ No ( )$_1$ Yes ( )$_8$ Don't know	
c. some other problem or disorder related to the urinary tract (specify) _____	( )$_0$ No ( )$_1$ Yes ( )$_8$ Don't know	
d. Some other problem or disorder related to the prostate (specify) _____	( )$_0$ No ( )$_1$ Yes ( )$_8$ Don't know	

I16. Have you ever had any prostate surgery?

( )$_0$ No **(Skip to I19)**
( )$_1$ Yes

I17. How many prostate surgeries have you had? _____

J18.	Year of surgery	Hospital name	City	State
a.				
b.				
c.				

I19. Were you ever treated by a doctor for a urinary tract infection since the age of 25?

( )$_0$ No **(Skip to I22)**
( )$_1$ Yes

I20. How old were you when your doctor first told you that you had a urinary tract infection?

___ ___ years

I21. How many times have you been diagnosed with a UTI?_____

I22. Have you had a vasectomy, that is a sterilization operation for men?

( )$_0$ No **(Skip to I24)**
( )$_1$ Yes

I23. How old were you when you had a vasectomy? __ __ years

I24. Were you circumcised? *Circumcision: The surgical removal of the foreskin of the penis.*

( )$_0$ No **(Skip to J)**
( )$_1$ Yes

I25. At what age were you circumcised?

( )$_1$ newborn
( )$_2$ other (specify in years) _____

PROSTATE HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

J. FAMILY MEDICAL HISTORY

J1. Has anyone in your family that is related to you by blood, ever been told he had Benign Prostatic Hyperplasia or an enlarged prostate? Include your sons, grandsons, father, paternal grandfather, maternal grandfather and brothers.

( )$_0$ No ( )$_1$ Yes

J2. If yes, at what age was it diagnosed?

Relative		Age at diagnosis (approximately) DK= 888
a	Brother(s) ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
b	Father ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
c	Son (s) ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
d	Maternal Grandfather ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
e	Paternal Grandfather ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
f	Other _____(specify) ( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	

J3. Has anyone in your family that is related to you by blood, ever been told he had prostate cancer? Include your sons, grandsons, father, paternal grandfather, maternal grandfather, brothers.

( )$_0$ No **(Skip to J5)** ( )$_1$ Yes

J4. If yes, at what age was it diagnosed?

Relative			Age at diagnosis (approximately) DK= 888
a	Brother(s)	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
b	Father	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
c	Son (s)	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
d	Maternal Grandfather	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
e	Paternal Grandfather	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
f	Other _____(specify)	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	

J5. Has any member of your family that is related to you by blood ever been told that she had breast cancer? Including your daughter, mother, sister, grandmothers.

( )$_0$ No **(Skip to J7)** ( )$_1$ Yes

J6. If yes, at what age was it diagnosed?

Relative			Age at diagnosis (approximately) DK= 888
a	Daughter	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
b	Mother	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
c	Sister	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
d	Maternal Grandmother	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
e	Paternal Grandmother	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	
f	Other _____(specify)	( )$_0$ No ( )$_1$ Yes ( )$_8$ DK	

J7. Have any members of your family that are related to you by blood ever been told that they had ovarian cancer? Please include your mother, daughter, and maternal and paternal grandmothers.

( )$_0$ No **(Skip to J9)** ( )$_1$ Yes

J8. If yes, at what age was it diagnosed?

Relative			Age at diagnosis (approximately) DK= 888
a	Daughter	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
b	Mother	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
c	Sister	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
d	Maternal Aunt	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
e	Paternal Grandmother	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
f	Other _____(specify)	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	

J9. Have any members of your family that are related to you by blood ever been told that they had endometrial cancer? Please include your mother, daughter, sisters and maternal and paternal grandmothers.

( )$_0$ No **(Skip to K)** ( )$_1$ Yes

J10. If yes, at what age was it diagnosed?

Relative			Age at diagnosis (approximately) DK= 888
a	Daughter	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
b	Mother	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
c	Sister(s)	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
d	Maternal Aunt	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
e	Paternal Grandmother	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	
f	Other _____(specify)	( )$_0$ No ( )$_1$ Yes ( )$_8$ D.K.	

FAMILY MEDICAL HISTORY ( )$_1$ Very Good ( )$_2$ Good ( )$_3$ Fair ( )$_4$ Poor

K. PHYSICAL ACTIVITY/EXERCISE

Now, we are going to ask you about your levels of physical activity at different times in your life.

	a. Last year	b. Age 13-19	c. 20s	d. 30s	e. 40s	f. 50s+
K1. Did you participate in any routine physical activity for at least 20 minutes at a time that either made you sweat or increased your heart rate?	0 No 1 Yes	0 No 1 Yes	0 No 1 Yes	0 No 1 Yes	0 No 1 Yes	0 No 1 Yes
K2. What intensity level was your usual activity?	1 Moderate 2 Vigorous	1 Moderate 2 Vigorous	1Moderate 2 Vigorous	1 Moderate 2 Vigorous	1 Moderate 2 Vigorous	1 Moderate 2 Vigorous
K3. How often did you participate in this physical activity?	1 <1x/week 2 1x/week 3 >1x/week	1 <1x/week 2 1x/week 3 >1x/week	1 <1x/week 2 1x/week 3 >1x/week	1 <1x/week 2 1x/week 3 >1x/week	1 <1x/week 2 1x/week 3 >1x/week	1 <1x/week 2 1x/week 3 >1x/week

PHYSICAL ACTIVITY ( )1 Very Good ( )2 Good ( )3 Fair ( )4 Poor

Section L (Sexual history) is self-administered, and the person will be given 20 min to complete this section.

SITE ID:______________________

L. SEXUAL HISTORY/HEALTH (self administered)

L1. At what age did you experience puberty (voice change, growth of pubic hair)? __ __ years

L2. How old were you when you first had sexual intercourse? __ __ years

	In your teens	In your 20's	In your 30's	In your 40's	In your 50's	In your 60's	In your 70's
L3.When you were (age group) with **how many different partners did you have** intercourse?	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ >40	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ >40	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ >40	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ >40	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ >40	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ >40	( )$_0$ 0 ( )$_1$ 1 ( )$_2$ 2 ( )$_3$ 3-4 ( )$_4$ 5-9 ( )$_5$ 10-19 ( )$_6$ 20-39 ( )$_7$ >40
L4.If you think back to when you were (age group), and you think about the period of time in that decade when you had sexual intercourse, **how often would you say you had sexual intercourse?** Fill in the box with the frequency and mark per month or per year.	[ ] times per ( ) month$_1$ ( ) year$_2$	[ ] times per ( ) month$_1$ ( ) year$_2$	[ ] times per ( ) month$_1$ ( ) year$_2$	[ ] times per ( ) month$_1$ ( ) year$_2$	[ ] times per ( ) month$_1$ ( ) year$_2$	[ ] times per ( ) month$_1$ ( ) year$_2$	[ ] times per ( ) month$_1$ ( ) year$_2$

SITE ID:_____

L5. How many live-born children have you fathered? Do not include any stepchildren, foster children, or adopted children. __ __ __ **(If zero, skip to L7)**

L6. How old were you when your first child was born? __ __ years

L7. Have you ever tried to conceive a child for one year or more without success? ( )$_0$ No ( )$_1$ Yes
(If NO, skip to L9)

L8. Did a doctor ever say that you had a problem that might be related to your difficulty in conceiving a child? If so, what was the problem?
( )$_0$ Low sperm count( )$_1$ Low sperm motility ( )$_2$ Impotence ( )$_3$ Other _____(specify)

L9. Have you ever used condoms (rubbers)? ( )$_0$ No **(If No, skip to L13)** ( )$_1$ Yes

L10 Not counting the times that you were trying to conceive a child, how often did you use condoms? ( )$_0$ Rarely ( )$_1$ Sometimes ( )$_2$ Always

L11. Before one year ago, did you usually use condoms (rubbers)? ( )$_0$ No ( )$_1$ Yes

L12. Not counting the past year, for how many years did you use condoms (rubbers)?
_____ YEARS

For the next question, please think about any sexually transmitted diseases that you may have contracted during your life.

L13.	Did a doctor ever tell you that you had:	Yes/No	How old were you when you were first diagnosed?	How many times altogether have you had the disease?
a.	Gonorrhea	( )$_0$ No ( )$_1$ Yes		
b.	Syphilis	( )$_0$ No ( )$_1$ Yes		
c.	Genital Warts	( )$_0$ No ( )$_1$ Yes		
d.	Genital Herpes	( )$_0$ No ( )$_1$ Yes		
e.	Other sexually transmitted disease _____(specify)	( )$_0$ No ( )$_1$ Yes		
f.	Other sexually transmitted disease _____(specify)	( )$_0$ No ( )$_1$ Yes		

This completes our interview. I would like to now take the samples and I want to thank you very much for the time you have spent in answering my questions today.

May we contact you again later if we need to clarify any of the information you have provided?
( )0 No ( )1 Yes

Time ended: ___ ___ : ___ ___ ( )1 AM
( )2 PM

M. ADMINISTRATIVE INFORMATION

M1. Date form completed ___ ___ / ___ ___ / ___ ___ ___ ___

M2. Name of interviewer _____/_____/_____

M3. Interviewer ID Number: ___ ___

M4. Interviewer's Signature: _____

N. INTERVIEWER REMARKS

N1. Interview was conducted:
( )1 In the clinic_____
( )2 General Clinical Research Center
( )3 Over the phone
( )4 Other (specify)_____

N2. Respondent's cooperation was:
( )1 Very good
( )2 Good
( )3 Fair
( )4 Poor

N3. The overall quality of the interview was:
( )1 Very good
( )2 Good
( )3 Fair
( )4 Poor

N4. Did any of the following occur during the interview?

a.	R did not know enough information regarding the topics.	( )0 No	( )1 Yes
b.	R did not want to be more specific.	( )0 No	( )1 Yes
c.	R did not understand or speak English well.	( )0 No	( )1 Yes
d.	R was upset or depressed.	( )0 No	( )1 Yes
e.	R had poor hearing or speech.	( )0 No	( )1 Yes
f.	R was confused by frequent interruptions.	( )0 No	( )1 Yes
g.	R was emotionally unstable.	( )0 No	( )1 Yes
h.	Others helped with the answers.	( )0 No	( )1 Yes

i.	R required a lot of probing	( )$_0$ No ( )$_1$ Yes
j.	Patient was reserved	( )$_0$ No ( )$_1$ Yes
k.	R was physically ill	( )$_0$ No ( )$_1$ Yes
l.	Other, (specify) _____	( )$_0$ No ( )$_1$ Yes

N5. Comments/Remarks:

_____
_____
_____
_____
_____
_____
_____

NATIONAL INSTITUTES OF HEALTH

Diet History Questionnaire

GENERAL INSTRUCTIONS

- **Answer each question as best you can. Estimate if you are not sure. A guess is better than leaving a blank.**
- **Use only a black ball-point pen. Do not use a pencil or felt-tip pen. Do not fold, staple, or tear the pages.**
- **Put an X in the box next to your answer.**
- **If you make any changes, cross out the incorrect answer and put an X in the box next to the correct answer. Also draw a circle around the correct answer.**
- **If you mark NEVER, NO, or DON'T KNOW for a question, please follow any arrows or instructions that direct you to the next question.**

BEFORE TURNING THE PAGE, PLEASE COMPLETE THE FOLLOWING QUESTIONS.

Today's date:

MONTH	DAY		YEAR
☐ Jan	|___|___|		☐ 2002
☐ Feb	☐0	☐0	☐ 2003
☐ Mar	☐1	☐1	☐ 2004
☐ Apr	☐2	☐2	☐ 2005
☐ May	☐3	☐3	☐ 2006
☐ Jun		☐4	
☐ Jul		☐5	
☐ Aug		☐6	
☐ Sep		☐7	
☐ Oct		☐8	
☐ Nov		☐9	
☐ Dec			

In what month were you born?

☐ Jan
☐ Feb
☐ Mar
☐ Apr
☐ May
☐ Jun
☐ Jul
☐ Aug
☐ Sep
☐ Oct
☐ Nov
☐ Dec

In what year were you born?

19 |___|___|

☐0	☐0
☐1	☐1
☐2	☐2
☐3	☐3
☐4	☐4
☐5	☐5
☐6	☐6
☐7	☐7
☐8	☐8
☐9	☐9

Are you male or female?

☐Male
☐Female

BAR CODE LABEL OR SUBJECT ID HERE

|___|___|___|___|___|___|___|___|

1. Over the past 12 months, how often did you drink **tomato juice** or **vegetable juice**?

☐ NEVER (GO TO QUESTION 2)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

1a. Each time you drank **tomato juice** or **vegetable juice**, how much did you usually drink?

☐ Less than ¾ cup (6 ounces)
☐ ¾ to 1¼ cups (6 to 10 ounces)
☐ More than 1¼ cups (10 ounces)

2. Over the past 12 months, how often did you drink **orange juice** or **grapefruit juice?**

☐ NEVER (GO TO QUESTION 3)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

2a. Each time you drank **orange juice** or **grapefruit juice**, how much did you usually drink?

☐ Less than ¾ cup (6 ounces)
☐ ¾ to 1¼ cups (6 to 10 ounces)
☐ More than 1¼ cups (10 ounces)

3. Over the past 12 months, how often did you drink **other 100% fruit juice** or **100% fruit juice mixtures** (such as apple, grape, pineapple, or others)?

☐ NEVER (GO TO QUESTION 4)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

3a. Each time you drank **other fruit juice** or **fruit juice mixtures**, how much did you usually drink?

☐ Less than ¾ cup (6 ounces)
☐ ¾ to 1½ cups (6 to 12 ounces)
☐ More than 1½ cups (12 ounces)

Question 4 appears in the next column

Over the past 12 months…

4. How often did you drink other **fruit drinks** (such as cranberry cocktail, Hi-C, lemonade, or Kool-Aid, diet or regular)?

☐ NEVER (GO TO QUESTION 5)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

4a. Each time you drank **fruit drinks**, how much did you usually drink?

☐ Less than 1 cup (8 ounces)
☐ 1 to 2 cups (8 to 16 ounces)
☐ More than 2 cups (16 ounces)

4b. How often were your fruit drinks **diet** or **sugar-free drinks**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

5. How often did you drink **milk as a beverage** (NOT in coffee, NOT in cereal)? (Please include chocolate milk and hot chocolate.)

☐ NEVER (GO TO QUESTION 6)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

5a. Each time you drank **milk as a beverage**, how much did you usually drink?

☐ Less than 1 cup (8 ounces)
☐ 1 to 1½ cups (8 to 12 ounces)
☐ More than 1½ cups (12 ounces)

5b. What kind of **milk** did you usually drink?

☐ Whole milk
☐ 2% fat milk
☐ 1 % fat milk
☐ Skim, nonfat, or ½% fat milk
☐ Soy milk
☐ Rice milk
☐ Other

Question 6 appears on the next page

Over the <u>past 12 months</u>…

6. How often did you drink **meal replacement, energy, or high-protein beverages** such as Instant Breakfast, Ensure, Slimfast, Sustacal or others?

☐ NEVER (GO TO QUESTION 7)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

6a. Each time you drank **meal replacement beverages**, how much did you usually drink?

☐ Less than 1 cup (8 ounces)
☐ 1 to 1½ cups (8 to 12 ounces)
☐ More than 1½ cups (12 ounces)

7. Over the past 12 months, did you drink **soft drinks, soda,** or **pop**?

☐ NO (GO TO QUESTION 8)

☐ YES

7a. How often did you drink **soft drinks, soda,** or **pop IN THE SUMMER**?

☐ NEVER

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

7b. How often did you drink **soft drinks, soda,** or **pop DURING THE REST OF THE YEAR**?

☐ NEVER

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

7c. Each time you drank **soft drinks, soda,** or **pop,** how much did you usually drink?

☐ Less than 12 ounces or less than 1 can or bottle
☐ 12 to 16 ounces or 1 can or bottle
☐ More than 16 ounces or more than 1 can or bottle

7d. How often were these soft drinks, soda, or pop **diet** or **sugar-free**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

7e. How often were these soft drinks, soda, or pop **caffeine-free**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

8. Over the <u>past 12 months</u>, did you drink **beer**?

☐ NO (GO TO QUESTION 9)

☐ YES

8a. How often did you drink **beer IN THE SUMMER**?

☐ NEVER

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

8b. How often did you drink **beer DURING THE REST OF THE YEAR**?

☐ NEVER

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

8c. Each time you drank **beer**, how much did you usually drink?

☐ Less than a 12-ounce can or bottle
☐ 1 to 3 12-ounce cans or bottles
☐ More than 3 12-ounce cans or bottles

Question 8 appears in the next column

Question 9 appears on the next page

Over the <u>past 12 months</u>…

9. How often did you drink **wine** or **wine coolers**?

☐ NEVER (GO TO QUESTION 10)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

9a. Each time you drank **wine** or **wine coolers,** how much did you usually drink?

☐ Less than 5 ounces or less than 1 glass
☐ 5 to 12 ounces or 1 to 2 glasses
☐ More than 12 ounces or more than 2 glasses

10. How often did you drink **liquor** or **mixed drinks**?

☐ NEVER (GO TO QUESTION 11)

☐ 1 time per month or less
☐ 2–3 times per month
☐ 1–2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

10a. Each time you drank **liquor** or **mixed drinks,** how much did you usually drink?

☐ Less than 1 shot of liquor
☐ 1 to 3 shots of liquor
☐ More than 3 shots of liquor

11. Over the <u>past 12 months</u>, did you eat **oatmeal, grits,** or **other cooked cereal?**

☐ NO (GO TO QUESTION 12)

☐ YES

11a. How often did you eat **oatmeal, grits,** or **other cooked cereal IN THE WINTER**?

☐ NEVER

☐ 1–6 times per winter
☐ 7–11 times per winter
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

11b. How often did you eat **oatmeal, grits,** or **other cooked cereal DURING THE REST OF THE YEAR**?

☐ NEVER

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

11c. Each time you ate **oatmeal, grits,** or **other cooked cereal,** how much did you usually eat?

☐ Less than ¾ cup
☐ ¾ to 1¼ cups
☐ More than 1¼ cups

12. How often did you eat **cold cereal**?

☐ NEVER (GO TO QUESTION 13)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

12a. Each time you ate **cold cereal**, how much did you usually eat?

☐ Less than 1 cup
☐ 1 to 2½ cups
☐ More than 2½ cups

12b. How often was the cold cereal you ate **Total, Product 19,** or **Right Start**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

12c. How often was the cold cereal you ate **All Bran, Fiber One, 100% Bran,** or **Bran Buds?**

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 12 appears in the next column

Question 13 appears on the next page

Over the <u>past 12 months</u>…

12d. How often was the cold cereal you ate **some other bran** or **fiber cereal** (such as Cheerios, Shredded Wheat, Raisin Bran, Bran Flakes, Grape-Nuts, Granola, Wheaties, or Healthy Choice)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

12e. How often was the cold cereal you ate any **other type of cold cereal** (such as Corn Flakes, Rice Krispies, Frosted Flakes, Special K, Froot Loops, Cap'n Crunch, or others)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

12f. Was **milk** added to your cold cereal?

- ☐ NO (GO TO QUESTION 13)
- ☐ YES

12g. What kind of **milk** was usually added?

- ☐ Whole milk
- ☐ 2% fat milk
- ☐ 1% fat milk
- ☐ Skim, nonfat, or ½% fat milk
- ☐ Soy milk
- ☐ Rice milk
- ☐ Other

12h. Each time **milk was added to your cold cereal**, how much was usually added?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

13. How often did you eat **applesauce**?

- ☐ NEVER (GO TO QUESTION 14)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

Question 14 appears in the next column

13a. Each time you ate **applesauce**, how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

14. How often did you eat **apples**?

- ☐ NEVER (GO TO QUESTION 15)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

14a. Each time you ate **apples**, how many did you usually eat?

- ☐ Less than 1 apple
- ☐ 1 apple
- ☐ More than 1 apple

15. How often did you eat **pears** (fresh, canned, or frozen)?

- ☐ NEVER (GO TO QUESTION 16)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

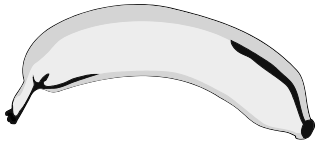
15a. Each time you ate **pears**, how many did you usually eat?

- ☐ Less than 1 pear
- ☐ 1 pear
- ☐ More than 1 pear

16. How often did you eat **bananas?**

- ☐ NEVER (GO TO QUESTION 17)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day



Question 17 appears on the next page

Over the <u>past 12 months</u>…

16a. Each time you ate **bananas**, how many did you usually eat?

- ☐ Less than 1 banana
- ☐ 1 banana
- ☐ More than 1 banana

17. How often did you eat **dried fruit**, such as prunes or raisins (not including dried apricots)?

- ☐ NEVER (GO TO QUESTION 18)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

17a. Each time you ate **dried fruit**, how much did you usually eat (not including dried apricots)?

- ☐ Less than 2 tablespoons
- ☐ 2 to 5 tablespoons
- ☐ More than 5 tablespoons

18. Over the <u>past 12 months</u>, did you eat **peaches, nectarines,** or **plums**?

- ☐ NO (GO TO QUESTION 19)
- ☐ YES

18a. How often did you eat **fresh peaches**, **nectarines**, or **plums WHEN IN SEASON**?

- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

18b. How often did you eat **peaches, nectarines,** or **plums** (fresh, canned, or frozen) **DURING THE REST OF THE YEAR**?

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 19 appears in the next column

18c. Each time you ate **peaches, nectarines,** or **plums,** how much did you usually eat?

- ☐ Less than 1 fruit or less than ½ cup
- ☐ 1 to 2 fruits or ½ to ¾ cup
- ☐ More than 2 fruits or more than ¾ cup

19. How often did you eat **grapes**?

- ☐ NEVER (GO TO QUESTION 20)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

19a. Each time you ate **grapes**, how much did you usually eat?

- ☐ Less than ½ cup or less than 10 grapes
- ☐ ½ to 1 cup or 10 to 30 grapes
- ☐ More than 1 cup or more than 30 grapes

20. Over the <u>past 12 months</u>, did you eat **cantaloupe**?

- ☐ NO (GO TO QUESTION 21)
- ☐ YES

20a. How often did you eat **fresh cantaloupe WHEN IN SEASON**?

- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

20b. How often did you eat **fresh** or **frozen cantaloupe DURING THE REST OF THE YEAR**?

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 21 appears on the next page

Over the <u>past 12 months</u>…

20c. Each time you ate **cantaloupe**, how much did you usually eat?

- ☐ Less than ¼ melon or less than ½ cup
- ☐ ¼ melon or ½ to 1 cup
- ☐ More than ¼ melon or more than 1 cup

21. Over the <u>past 12 months</u>, did you eat **melon, other than cantaloupe** (such as watermelon or honeydew)?

- ☐ NO (GO TO QUESTION 22)
- ☐ YES

21a. How often did you eat **fresh melon, other than cantaloupe** (such as watermelon or honeydew) **WHEN IN SEASON**?

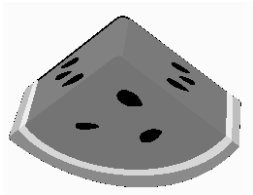
- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

21b. How often did you eat **fresh** or **frozen melon, other than cantaloupe** (such as watermelon or honeydew) **DURING THE REST OF THE YEAR**?

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

21c. Each time you ate **melon other than cantaloupe**, how much did you usually eat?

- ☐ Less than ½ cup or 1 small wedge
- ☐ ½ to 2 cups or 1 medium wedge
- ☐ More than 2 cups or 1 large wedge



Question 22 appears in the next column

22. Over the <u>past 12 months</u>, did you eat **strawberries**?

- ☐ NO (GO TO QUESTION 23)
- ☐ YES

22a. How often did you eat **fresh strawberries WHEN IN SEASON**?

- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

22b. How often did you eat **fresh** or **frozen strawberries DURING THE REST OF THE YEAR**?

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

22c. Each time you ate **strawberries**, how much did you usually eat?

- ☐ Less than ¼ cup or less than 3 berries
- ☐ ¼ to ¾ cup or 3 to 8 berries
- ☐ More than ¾ cup or more than 8 berries

23. Over the <u>past 12 months</u>, did you eat **oranges, tangerines,** or **tangelos**?

- ☐ NO (GO TO QUESTION 24)
- ☐ YES

23a. How often did you eat **fresh oranges, tangerines,** or **tangelos WHEN IN SEASON**?

- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 24 appears on the next page

Over the <u>past 12 months</u>…

23b. How often did you eat **oranges, tangerines,** or **tangelos** (fresh or canned) **DURING THE REST OF THE YEAR**?

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

23c. Each time you ate **oranges, tangerines,** or **tangelos,** how many did you usually eat?

- ☐ Less than 1 fruit
- ☐ 1 fruit
- ☐ More than 1 fruit

24. Over the <u>past 12 months</u>, did you eat **grapefruit**?

- ☐ NO (GO TO QUESTION 25)
- ☐ YES

24a. How often did you eat **fresh grapefruit WHEN IN SEASON**?

- ☐ NEVER
- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

24b. How often did you eat **grapefruit** (fresh or canned) **DURING THE REST OF THE YEAR?**

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

24c. Each time you ate **grapefruit**, how much did you usually eat?

- ☐ Less than ½ grapefruit
- ☐ ½ grapefruit
- ☐ More than ½ grapefruit

Question 25 appears in the next column

25. How often did you eat **other kinds of fruit**?

- ☐ NEVER (GO TO QUESTION 26)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

25a. Each time you ate **other kinds of fruit**, how much did you usually eat?

- ☐ Less than ¼ cup
- ☐ ¼ to ¾ cup
- ☐ More than ¾ cup

26. How often did you eat **COOKED greens** (such as spinach, turnip, collard, mustard, chard, or kale)?

- ☐ NEVER (GO TO QUESTION 27)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

26a. Each time you ate **COOKED greens**, how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

27. How often did you eat **RAW greens** (such as spinach, turnip, collard, mustard, chard, or kale)? (*We will ask about lettuce later.)*

- ☐ NEVER (GO TO QUESTION 28)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

27a. Each time you ate **RAW greens**, how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

Question 28 appears on the next page

Over the <u>past 12 months</u>…

28. How often did you eat **coleslaw**?

☐ NEVER (GO TO QUESTION 29)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

28a. Each time you ate **coleslaw**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to ¾ cup
☐ More than ¾ cup

29. How often did you eat **sauerkraut** or **cabbage** (other than coleslaw)?

☐ NEVER (GO TO QUESTION 30)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

29a. Each time you ate **sauerkraut** or **cabbage**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to 1 cup
☐ More than 1 cup

30. How often did you eat **carrots** (fresh, canned, or frozen)?

☐ NEVER (GO TO QUESTION 31)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

30a. Each time you ate **carrots**, how much did you usually eat?

☐ Less than ¼ cup or less than 2 baby carrots
☐ ¼ to ½ cup or 2 to 5 baby carrots
☐ More than ½ cup or more than 5 baby carrots

Question 31 appears in the next column

31. How often did you eat **string beans** or **green beans** (fresh, canned, or frozen)?

☐ NEVER (GO TO QUESTION 32)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

31a. Each time you ate **string beans** or **green beans**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 cup
☐ More than 1 cup

32. How often did you eat **peas** (fresh, canned, or frozen)?

☐ NEVER (GO TO QUESTION 33)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

32a. Each time you ate **peas**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to ¾ cup
☐ More than ¾ cup

33. Over the <u>past 12 months</u>, did you eat **corn**?

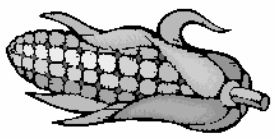
☐ NO (GO TO QUESTION 34)

☐ YES

33a. How often did you eat **fresh corn WHEN IN SEASON**?

☐ NEVER

☐ 1–6 times per season
☐ 7–11 times per season
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day



Question 34 appears on the next page

Over the <u>past 12 months</u>…

33b. How often did you eat **corn** (fresh, canned, or frozen) **DURING THE REST OF THE YEAR**?

☐ NEVER

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

33c. Each time you ate **corn**, how much did you usually eat?

☐ Less than 1 ear or less than ½ cup
☐ 1 ear or ½ to 1 cup
☐ More than 1 ear or more than 1 cup

34. Over the <u>past 12 months</u>, how often did you eat **broccoli** (fresh or frozen)?

☐ NEVER (GO TO QUESTION 35)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

34a. Each time you ate **broccoli**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to 1 cup
☐ More than 1 cup

35. How often did you eat **cauliflower** or **Brussels sprouts** (fresh or frozen)?

☐ NEVER (GO TO QUESTION 36)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

35a. Each time you ate **cauliflower** or **Brussels sprouts**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to ½ cup
☐ More than ½ cup

Question 36 appears in the next column

36. How often did you eat mixed **vegetables**?

☐ NEVER (GO TO QUESTION 37)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

36a. Each time you ate **mixed vegetables**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 cup
☐ More than 1 cup

37. How often did you eat **onions**?

☐ NEVER (GO TO QUESTION 38)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

37a. Each time you ate **onions**, how much did you usually eat?

☐ Less than 1 slice or less than 1 tablespoon
☐ 1 slice or 1 to 4 tablespoons
☐ More than 1 slice or more than 4 tablespoons

38. Now think about all the **cooked vegetables** you ate in the <u>past 12 months</u> and how they were prepared. How often were your vegetables **COOKED WITH** some sort of **fat**, including oil spray? *(Please do not include potatoes.)*

☐ NEVER (GO TO QUESTION 39)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

Question 39 appears on the next page

Over the <u>past 12 months</u>…

38a. Which fats were usually added to your vegetables **DURING COOKING**? *(Please do not include potatoes. **Mark all that apply.**)*

- ☐ Margarine (including low-fat)
- ☐ Butter (including low-fat)
- ☐ Lard, fatback, or bacon fat
- ☐ Olive oil
- ☐ Corn oil
- ☐ Canola or rapeseed oil
- ☐ Oil spray, such as Pam or others
- ☐ Other kinds of oils
- ☐ None of the above

39. Now, thinking again about all the **cooked vegetables** you ate in the <u>past 12 months</u>, how often was some sort of fat, sauce, or dressing added **AFTER COOKING OR AT THE TABLE**? *(Please do not include potatoes.)*

☐ NEVER (GO TO QUESTION 40)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1–2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 times per day
- ☐ 3 or more times per day

39a. Which fats, sauces, or dressings were usually added **AFTER COOKING OR AT THE TABLE**? *(Please do not include potatoes. **Mark all that apply.**)*

- ☐ Margarine (including low-fat)
- ☐ Butter (including low-fat)
- ☐ Lard, fatback, or bacon fat
- ☐ Salad dressing
- ☐ Cheese sauce
- ☐ White sauce
- ☐ Other

39b. If margarine, butter, lard, fatback, or bacon fat was added to your cooked vegetables **AFTER COOKING OR AT THE TABLE**, how much did you usually add?

- ☐ Did not usually add these
- ☐ Less than 1 teaspoon
- ☐ 1 to 3 teaspoons
- ☐ More than 3 teaspoons

39c. If salad dressing, cheese sauce, or white sauce was added to your cooked vegetables **AFTER COOKING OR AT THE TABLE**, how much did you usually add?

- ☐ Did not usually add these
- ☐ Less than 1 tablespoon
- ☐ 1 to 3 tablespoons
- ☐ More than 3 tablespoons

Question 40 appears in the next column

40. Over the <u>past 12 months</u>, how often did you eat **sweet peppers** (green, red, or yellow)?

☐ NEVER (GO TO QUESTION 41)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

40a. Each time you ate **sweet peppers**, how much did you usually eat?

- ☐ Less than ⅛ pepper
- ☐ ⅛ to ¼ pepper
- ☐ More than ¼ pepper

41. Over the <u>past 12 months</u>, did you eat **fresh tomatoes** (including those in salads)?

☐ NO (GO TO QUESTION 42)

☐ YES

41a. How often did you eat **fresh tomatoes** (including those in salads) **WHEN IN SEASON**?

☐ NEVER

- ☐ 1–6 times per season
- ☐ 7–11 times per season
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

41b. How often did you eat **fresh tomatoes** (including those in salads) **DURING THE REST OF THE YEAR**?

☐ NEVER

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

41c. Each time you ate **fresh tomatoes**, how much did you usually eat?

- ☐ Less than ¼ tomato
- ☐ ¼ to ½ tomato
- ☐ More than ½ tomato

Question 42 appears on the next page

Over the <u>past 12 months</u>…

42. How often did you eat **lettuce salads** (with or without other vegetables)?

☐ NEVER (GO TO QUESTION 43)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

42a. Each time you ate **lettuce salads**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to 1 /4 cups
☐ More than 1 /4 cups

43. How often did you eat **salad dressing** (including low-fat) on salads?

☐ NEVER (GO TO QUESTION 44)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

43a. Each time you ate **salad dressing** on salads, how much did you usually eat?

☐ Less than 2 tablespoons
☐ 2 to 4 tablespoons
☐ More than 4 tablespoons

44. How often did you eat **sweet potatoes** or **yams**?

☐ NEVER (GO TO QUESTION 45)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

44a. Each time you ate **sweet potatoes** or **yams**, how much did you usually eat?

☐ 1 small potato or less than ¼ cup
☐ 1 medium potato or ¼ to ¾ cup
☐ 1 large potato or more than ¾ cup

45. How often did you eat **French fries, home fries, hash browned potatoes**, or **tater tots**?

☐ NEVER (GO TO QUESTION 46)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

45a. Each time you ate **French fries**, **home fries**, **hash browned potatoes**, or **tater tots** how much did you usually eat?

☐ Less than 10 fries or less than ½ cup
☐ 10 to 25 fries or ½ to 1 cup
☐ More than 25 fries or more than 1 cup

46. How often did you eat **potato salad**?

☐ NEVER (GO TO QUESTION 47)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

46a. Each time you ate **potato salad**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 cup
☐ More than 1 cup

47. How often did you eat **baked, boiled,** or **mashed potatoes**?

☐ NEVER (GO TO QUESTION 48)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

47a. Each time you ate **baked, boiled,** or **mashed potatoes**, how much did you usually eat?

☐ 1 small potato or less than ½ cup
☐ 1 medium potato or ½ to 1 cup
☐ 1 large potato or more than 1 cup

Question 45 appears in the next column

Question 48 appears on the next page

Over the <u>past 12 months</u>…

47b. How often was **sour cream** (including low-fat) added to your potatoes, **EITHER IN COOKING OR AT THE TABLE**?

- ☐ Almost never or never (GO TO QUESTION 47d)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

47c. Each time **sour cream** was added to your potatoes, how much was usually added?

- ☐ Less than 1 tablespoon
- ☐ 1 to 3 tablespoons
- ☐ More than 3 tablespoons

47d. How often was **margarine** (including low-fat) added to your potatoes, **EITHER IN COOKING OR AT THE TABLE**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

47e. How often was **butter** (including low-fat) added to your potatoes, **EITHER IN COOKING OR AT THE TABLE**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

47f. Each time **margarine** or **butter** was added to your potatoes, how much was usually added?

- ☐ Never added
- ☐ Less than 1 teaspoon
- ☐ 1 to 3 teaspoons
- ☐ More than 3 teaspoons

47g. How often was cheese or cheese sauce added to your potatoes, EITHER IN COOKING OR AT THE TABLE?

- ☐ Almost never or never (GO TO QUESTION 48)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

47h. Each time **cheese** or **cheese sauce** was added to your potatoes, how much was usually added?

- ☐ Less than 1 tablespoon
- ☐ 1 to 3 tablespoons
- ☐ More than 3 tablespoons

48. How often did you eat **salsa**?

- ☐ NEVER (GO TO QUESTION 49)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

48a. Each time you ate **salsa**, how much did you usually eat?

- ☐ Less than 1 tablespoon
- ☐ 1 to 5 tablespoons
- ☐ More than 5 tablespoons

49. How often did you eat **catsup**?

- ☐ NEVER (GO TO QUESTION 50)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

49a. Each time you ate **catsup**, how much did you usually eat?

- ☐ Less than 1 teaspoon
- ☐ 1 to 6 teaspoons
- ☐ More than 6 teaspoons

50. How often did you eat **stuffing, dressing,** or **dumplings**?

- ☐ NEVER (GO TO QUESTION 51)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

50a. Each time you ate **stuffing, dressing,** or **dumplings**, how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

Question 48 appears in the next column

Question 51 appears on the next page

Over the <u>past 12 months</u>…

51. How often did you eat **chili**?

☐ NEVER (GO TO QUESTION 52)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

51a. Each time you ate **chili**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 3/4 cups
☐ More than 1 3/4 cups

52. How often did you eat **Mexican foods** (such as tacos, tostados, burritos, tamales, fajitas, enchiladas, quesadillas, and chimichangas)?

☐ NEVER (GO TO QUESTION 53)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

52a. Each time you ate **Mexican foods**, how much did you usually eat?

☐ Less than 1 taco, burrito, etc.
☐ 1 to 2 tacos, burritos, etc.
☐ More than 2 tacos, burritos, etc.

53. How often did you eat **cooked dried beans** (such as baked beans, pintos, kidney, blackeyed peas, lima, lentils, soybeans, or refried beans)? *(Please don't include bean soups or chili.)*

☐ NEVER (GO TO QUESTION 54)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

53a. Each time you ate **beans**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 cup
☐ More than 1 cup

53b. How often were the beans you ate **refried beans, beans prepared with any type of fat**, or **with meat added**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

54. How often did you eat **other kinds of vegetables**?

☐ NEVER (GO TO QUESTION 55)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

54a. Each time you ate **other kinds of vegetables**, how much did you usually eat?

☐ Less than ¼ cup
☐ ¼ to ½ cup
☐ More than ½ cup

55. How often did you eat **rice** or **other cooked grains** (such as bulgur, cracked wheat, or millet)?

☐ NEVER (GO TO QUESTION 56)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

55a. Each time you ate **rice** or **other cooked grains**, how much did you usually eat?

☐ Less than ½ cup
☐ ½ to 1 /2 cups
☐ More than 1 /2 cups

55b. How often was **butter, margarine**, or **oil** added to your rice **IN COOKING OR AT THE TABLE**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 54 appears in the next column

Question 56 appears on the next page

Over the <u>past 12 months</u>…

56. How often did you eat **pancakes, waffles,** or **French toast**?

- ☐ NEVER (GO TO QUESTION 57)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

56a. Each time you ate **pancakes, waffles,** or **French toast**, how much did you usually eat?

- ☐ Less than 1 medium piece
- ☐ 1 to 3 medium pieces
- ☐ More than 3 medium pieces

56b. How often was **margarine** (including low-fat) added to your pancakes, waffles, or French toast **AFTER COOKING OR AT THE TABLE**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

56c. How often was **butter** (including low-fat) added to your pancakes, waffles, or French toast **AFTER COOKING OR AT THE TABLE**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

56d. Each time **margarine** or **butter** was added to your pancakes, waffles, or French toast, how much was usually added?

- ☐ Never added
- ☐ Less than 1 teaspoon
- ☐ 1 to 3 teaspoons
- ☐ More than 3 teaspoons

56e. How often was **syrup** added to your pancakes, waffles, or French toast?

- ☐ Almost never or never (GO TO QUESTION 57)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 57 appears in the next column

56f. Each time **syrup** was added to your pancakes, waffles, or French toast, how much was usually added?

- ☐ Less than 1 tablespoon
- ☐ 1 to 4 tablespoons
- ☐ More than 4 tablespoons

57. How often did you eat **lasagna, stuffed shells, stuffed manicotti, ravioli,** or **tortellini**? *(Please do not include spaghetti or other pasta.)*

- ☐ NEVER (GO TO QUESTION 58)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

57a. Each time you ate **lasagna, stuffed shells, stuffed manicotti, ravioli,** or **tortellini**, how much did you usually eat?

- ☐ Less than 1 cup
- ☐ 1 to 2 cups
- ☐ More than 2 cups

58. How often did you eat **macaroni and cheese**?

- ☐ NEVER (GO TO QUESTION 59)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

58a. Each time you ate **macaroni and cheese**, how much did you usually eat?

- ☐ Less than 1 cup
- ☐ 1 to 1 /2 cups
- ☐ More than 1 /2 cups

59. How often did you eat **pasta salad** or **macaroni salad**?

- ☐ NEVER (GO TO QUESTION 60)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 60 appears on the next page

Over the <u>past 12 months</u>…

59a. Each time you ate **pasta salad** or **macaroni salad**, how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

60. Other than the pastas listed in Questions 57, 58, and 59, how often did you eat **pasta, spaghetti**, or **other noodles**?

- ☐ NEVER (GO TO QUESTION 61)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

60a. Each time you ate **pasta, spaghetti**, **or other noodles**, how much did you usually eat?

- ☐ Less than 1 cup
- ☐ 1 to 3 cups
- ☐ More than 3 cups

60b. How often did you eat your pasta, spaghetti, or other noodles with **tomato sauce** or **spaghetti sauce made WITH meat**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

60c. How often did you eat your pasta, spaghetti, or other noodles with **tomato sauce** or **spaghetti sauce made WITHOUT meat**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

60d. How often did you eat your pasta, spaghetti, or other noodles with **margarine, butter, oil,** or **cream sauce**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 61 appears in the next column

61. How often did you eat **bagels** or **English muffins**?

- ☐ NEVER (GO TO INTRODUCTION TO QUESTION 62)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

61a. Each time you ate **bagels** or **English muffins**, how many did you usually eat?

- ☐ Less than 1 bagel or English muffin
- ☐ 1 bagel or English muffin
- ☐ More than 1 bagel or English muffin

61b. How often was **margarine** (including low-fat) added to your bagels or English muffins?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

61c. How often was **butter** (including low-fat) added to your bagels or English muffins?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

61d. Each time **margarine** or **butter** was added to your bagels or English muffins, how much was usually added?

- ☐ Never added
- ☐ Less than 1 teaspoon
- ☐ 1 to 2 teaspoons
- ☐ More than 2 teaspoons

61e. How often was **cream cheese** (including low-fat) spread on your bagels or English muffins?

- ☐ Almost never or never (GO TO INTRODUCTION TO QUESTION 62)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Introduction to Question 62 appears on the next page

Over the <u>past 12 months</u>…

61f. Each time **cream cheese** was added to your bagels or English muffins, how much was usually added?

- ☐ Less than 1 tablespoon
- ☐ 1 to 2 tablespoons
- ☐ More than 2 tablespoons

The next questions ask about your intake of breads other than bagels or English muffins. First, we will ask about bread you ate as part of sandwiches only. Then we will ask about all other bread you ate.

62. How often did you eat **breads** or **rolls AS PART OF SANDWICHES** (including burger and hot dog rolls)?

- ☐ NEVER (GO TO QUESTION 63)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

62a. Each time you ate **breads** or **rolls AS PART OF SANDWICHES**, how many did you usually eat?

- ☐ 1 slice or ½ roll
- ☐ 2 slices or 1 roll
- ☐ More than 2 slices or more than 1 roll

62b. How often were the breads or rolls that you used for your sandwiches **white bread** (including burger and hot dog rolls)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

62c. How often was **mayonnaise** or **mayonnaise-type dressing** (including low-fat) added to your sandwich bread or rolls?

- ☐ Almost never or never (GO TO QUESTION 62e)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

62d. Each time **mayonnaise** or **mayonnaise-type dressing** was added to your sandwich breads or rolls, how much was usually added?

- ☐ Less than 1 teaspoon
- ☐ 1 to 3 teaspoons
- ☐ More than 3 teaspoons

62e. How often was **margarine** (including low-fat) added to your sandwich bread or rolls?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

62f. How often was **butter** (including low-fat) added to your sandwich bread or rolls?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

62g. Each time **margarine** or **butter** was added to your sandwich breads or rolls, how much was usually added?

- ☐ Never added
- ☐ Less than 1 teaspoon
- ☐ 1 to 2 teaspoons
- ☐ More than 2 teaspoons

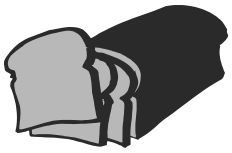
63. How often did you eat **breads** or **dinner rolls, NOT AS PART OF SANDWICHES**?

- ☐ NEVER (GO TO QUESTION 64)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

63a. Each time you ate **breads** or **dinner rolls, NOT AS PART OF SANDWICHES**, how much did you usually eat?

- ☐ 1 slice or 1 dinner roll
- ☐ 2 slices or 2 dinner rolls
- ☐ More than 2 slices or 2 dinner rolls



Question 62e appears in the next column

Question 63 appears in the next column

Question 64 appears on the next page

Over the <u>past 12 months</u>…

63b. How often were the breads or rolls you ate **white bread**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

63c. How often was **margarine** (including low-fat) added to your breads or rolls?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

63d. How often was **butter** (including low-fat) added to your breads or rolls?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

63e. Each time **margarine** or **butter** was added to your breads or rolls, how much was usually added?

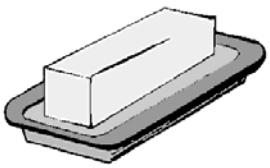
- ☐ Never added
- ☐ Less than 1 teaspoon
- ☐ 1 to 2 teaspoons
- ☐ More than 2 teaspoons

63f. How often was **cream cheese** (including low-fat) added to your breads or rolls?

- ☐ Almost never or never (GO TO QUESTION 64)
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

63g. Each time **cream cheese** was added to your breads or rolls, how much was usually added?

- ☐ Less than 1 tablespoon
- ☐ 1 to 2 tablespoons
- ☐ More than 2 tablespoons



Question 64 appears in the next column

64. How often did you eat **jam, jelly,** or **honey** on bagels, muffins, bread, rolls, or crackers?

- ☐ NEVER (GO TO QUESTION 65)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

64a. Each time you ate **jam, jelly,** or **honey**, how much did you usually eat?

- ☐ Less than 1 teaspoon
- ☐ 1 to 3 teaspoons
- ☐ More than 3 teaspoons

65. How often did you eat **peanut butter** or **other nut butter**?

- ☐ NEVER (GO TO QUESTION 66)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

65a. Each time you ate **peanut butter** or **other nut butter**, how much did you usually eat?

- ☐ Less than 1 tablespoon
- ☐ 1 to 2 tablespoons
- ☐ More than 2 tablespoons

66. How often did you eat **roast beef** or **steak IN SANDWICHES**?

- ☐ NEVER (GO TO QUESTION 67)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

66a. Each time you ate **roast beef** or **steak IN SANDWICHES**, how much did you usually eat?

- ☐ Less than 1 slice or less than 2 ounces
- ☐ 1 to 2 slices or 2 to 4 ounces
- ☐ More than 2 slices or more than 4 ounces

Question 67 appears on the next page

Over the <u>past 12 months</u>…

67. How often did you eat **turkey** or **chicken COLD CUTS** (such as loaf, luncheon meat, turkey ham, turkey salami, or turkey pastrami)? (*We will ask about other turkey or chicken later.)*

☐ NEVER (GO TO QUESTION 68)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

67a. Each time you ate **turkey or chicken COLD CUTS**, how much did you usually eat?

☐ Less than 1 slice
☐ 1 to 3 slices
☐ More than 3 slices

68. How often did you eat **luncheon** or **deli-style ham**? *(We will ask about other ham later.)*

☐ NEVER (GO TO QUESTION 69)

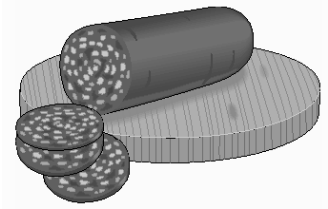
☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

68a. Each time you ate **luncheon** or **deli-style ham**, how much did you usually eat?

☐ Less than 1 slice
☐ 1 to 3 slices
☐ More than 3 slices

68b. How often was the luncheon or deli-style ham you ate **light**, **low-fat**, or **fat-free**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always



Question 69 appears in the next column

69. How often did you eat **other cold cuts** or **luncheon meats** (such as bologna, salami, corned beef, pastrami, or others, including low-fat)? *(Please do not include ham, turkey, or chicken cold cuts.)*

☐ NEVER (GO TO QUESTION 70)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

69a. Each time you ate **other cold cuts** or **luncheon meats**, how much did you usually eat?

☐ Less than 1 slice
☐ 1 to 3 slices
☐ More than 3 slices

69b. How often were the other cold cuts or luncheon meats you ate **light, low-fat**, or **fat-free cold cuts** or **luncheon meats**? *(Please do not include ham, turkey, or chicken cold cuts.)*

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

70. How often did you eat **canned tuna** (including in salads, sandwiches, or casseroles)?

☐ NEVER (GO TO QUESTION 71)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

70a. Each time you ate **canned tuna**, how much did you usually eat?

☐ Less than ¼ cup or less than 2 ounces
☐ ¼ to ½ cup or 2 to 3 ounces
☐ More than ½ cup or more than 3 ounces

70b. How often was the canned tuna you ate **water-packed tuna**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 71 appears on the next page

Over the <u>past 12 months</u>…

70c. How often was the canned tuna you ate **prepared with mayonnaise** or **other dressing** (including low-fat)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

71. How often did you eat **GROUND chicken** or **turkey**? *(We will ask about other chicken and turkey later.)*

- ☐ NEVER (GO TO QUESTION 72)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

71a. Each time you ate **GROUND chicken** or **turkey**, how much did you usually eat?

- ☐ Less than 2 ounces or less than ½ cup
- ☐ 2 to 4 ounces or ½ to 1 cup
- ☐ More than 4 ounces or more than 1 cup

72. How often did you eat **beef hamburgers** or **cheeseburgers**?

- ☐ NEVER (GO TO QUESTION 73)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

72a. Each time you ate **beef hamburgers** or **cheeseburgers**, how much did you usually eat?

- ☐ Less than 1 patty or less than 2 ounces
- ☐ 1 patty or 2 to 4 ounces
- ☐ More than 1 patty or more than 4 ounces

72b. How often were the beef hamburgers or cheeseburgers you ate made with **lean ground beef**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 73 appears in the next column

73. How often did you eat **ground beef in mixtures** (such as meatballs, casseroles, chili, or meatloaf)?

- ☐ NEVER (GO TO QUESTION 74)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

73a. Each time you ate **ground beef in mixtures**, how much did you usually eat?

- ☐ Less than 3 ounces or less than ½ cup
- ☐ 3 to 8 ounces or ½ to 1 cup
- ☐ More than 8 ounces or more than 1 cup

74. How often did you eat **hot dogs** or **frankfurters**? *(Please do not include sausages or vegetarian hot dogs.)*

- ☐ NEVER (GO TO QUESTION 75)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

74a. Each time you ate **hot dogs** or **frankfurters**, how many did you usually eat?

- ☐ Less than 1 hot dog
- ☐ 1 to 2 hot dogs
- ☐ More than 2 hot dogs

74b. How often were the hot dogs or frankfurters you ate **light** or **low-fat hot dogs**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 75 appears on the next page

Over the <u>past 12 months</u>…

75. How often did you eat beef mixtures such as **beef stew, beef pot pie, beef and noodles,** or **beef and vegetables**?

☐ NEVER (GO TO QUESTION 76)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

75a. Each time you ate **beef stew, beef pot pie, beef and noodles,** or **beef and vegetables,** how much did you usually eat?

☐ Less than 1 cup
☐ 1 to 2 cups
☐ More than 2 cups

76. How often did you eat **roast beef** or **pot roast**? *(Please do not include roast beef or pot roast in sandwiches.)*

☐ NEVER (GO TO QUESTION 77)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

76a. Each time you ate **roast beef** or **pot roast** (including in mixtures), how much did you usually eat?

☐ Less than 2 ounces
☐ 2 to 5 ounces
☐ More than 5 ounces

77. How often did you eat **steak** (beef)? (*Do not include steak in sandwiches*)

☐ NEVER (GO TO QUESTION 78)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

77a. Each time you ate **steak** (beef), how much did you usually eat?

☐ Less than 3 ounces
☐ 3 to 7 ounces
☐ More than 7 ounces

77b. How often was the steak you ate **lean steak**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

78. How often did you eat **pork** or **beef spareribs**?

☐ NEVER (GO TO QUESTION 79)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

78a. Each time you ate **pork** or **beef spareribs**, how much did you usually eat?

☐ Less than 4 ribs
☐ 4 to 12 ribs
☐ More than 12 ribs

79. How often did you eat **roast turkey, turkey cutlets,** or **turkey nuggets** (including in sandwiches)?

☐ NEVER (GO TO QUESTION 80)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

79a. Each time you ate **roast turkey, turkey cutlets,** or **turkey nuggets**, how much did you usually eat? *(Please note: 4 to 8 turkey nuggets = 3 ounces.)*

☐ Less than 2 ounces
☐ 2 to 4 ounces
☐ More than 4 ounces

80. How often did you eat **chicken** as part of **salads, sandwiches, casseroles, stews,** or **other mixtures**?

☐ NEVER (GO TO QUESTION 81)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

Question 78 appears in the next column

Question 81 appears on the next page

Over the <u>past 12 months</u>…

80a. Each time you ate **chicken** as part of **salads, sandwiches, casseroles, stews,** or **other mixtures,** how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 /2 cups
- ☐ More than 1 /2 cups

81. How often did you eat **baked, broiled, roasted, stewed,** or **fried chicken** (including nuggets)? *(Please do not include chicken in mixtures.)*

- ☐ NEVER (GO TO QUESTION 82)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

81a. Each time you ate **baked, broiled, roasted, stewed,** or **fried chicken** (including nuggets), how much did you usually eat?

- ☐ Less than 2 drumsticks or wings, less than 1 breast or thigh, or less than 4 nuggets
- ☐ 2 drumsticks or wings, 1 breast or thigh, or 4 to 8 nuggets
- ☐ More than 2 drumsticks or wings, more than 1 breast or thigh, or more than 8 nuggets

81b. How often was the chicken you ate **fried chicken** (including deep fried) or **chicken nuggets**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

81c. How often was the chicken you ate **WHITE meat**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

81d. How often did you eat chicken **WITH skin**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 82 appears in the next column

82. How often did you eat **baked ham** or **ham steak**?

- ☐ NEVER (GO TO QUESTION 83)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

82a. Each time you ate **baked ham** or **ham steak**, how much did you usually eat?

- ☐ Less than 1 ounce
- ☐ 1 to 3 ounces
- ☐ More than 3 ounces

83. How often did you eat **pork** (including chops, roasts, and in mixed dishes)? *(Please do not include ham, ham steak, or sausage.)*

- ☐ NEVER (GO TO QUESTION 84)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

83a. Each time you ate **pork**, how much did you usually eat?

- ☐ Less than 2 ounces or less than 1 chop
- ☐ 2 to 5 ounces or 1 chop
- ☐ More than 5 ounces or more than 1 chop

84. How often did you eat **gravy** on meat, chicken, potatoes, rice, etc.?

- ☐ NEVER (GO TO QUESTION 85)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

84a. Each time you ate **gravy** on meat, chicken, potatoes, rice, etc., how much did you usually eat?

- ☐ Less than ⅛ cup
- ☐ ⅛ to ½ cup
- ☐ More than ½ cup

Question 85 appears on the next page

Over the <u>past 12 months</u>…

85. How often did you eat **liver** (all kinds) or **liverwurst**?

☐ NEVER (GO TO QUESTION 86)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

85a. Each time you ate **liver** or **liverwurst**, how much did you usually eat?

☐ Less than 1 ounce
☐ 1 to 4 ounces
☐ More than 4 ounces

86. How often did you eat **bacon** (including low-fat)?

☐ NEVER (GO TO QUESTION 87)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

86a. Each time you ate **bacon**, how much did you usually eat?

☐ Fewer than 2 slices
☐ 2 to 3 slices
☐ More than 3 slices

86b. How often was the bacon you ate **light, low-fat,** or **lean bacon**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

87. How often did you eat **sausage** (including low-fat)?

☐ NEVER (GO TO QUESTION 88)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

87a. Each time you ate **sausage**, how much did you usually eat?

☐ Less than 1 patty or 2 links
☐ 1 to 3 patties or 2 to 5 links
☐ More than 3 patties or 5 links

87b. How often was the sausage you ate **light, low-fat**, or **lean sausage**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

88. How often did you eat **fish sticks** or **fried fish** (including fried seafood or shellfish)?

☐ NEVER (GO TO QUESTION 89)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

88a. Each time you ate **fish sticks** or **fried fish**, how much did you usually eat?

☐ Less than 2 ounces or less than 1 fillet
☐ 2 to 7 ounces or 1 fillet
☐ More than 7 ounces or more than 1 fillet

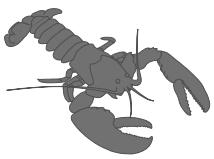
89. How often did you eat **fish** or **seafood that was NOT FRIED** (including shellfish)?

☐ NEVER (GO TO INTRODUCTION TO QUESTION 90)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

89a. Each time you ate eat **fish** or **seafood that was NOT FRIED**, how much did you usually eat?

☐ Less than 2 ounces or less than 1 fillet
☐ 2 to 5 ounces or 1 fillet
☐ More than 5 ounces or more than 1 fillet



Question 88 appears in the next column

Introduction to Question 90 appears on the next page

Over the <u>past 12 months</u>…

Now think about all the meat, poultry, and fish you ate in the <u>past 12 months</u> and how they were prepared.

90. How often was **oil, butter, margarine**, **or other fat** used to **FRY, SAUTE, BASTE, OR MARINATE** any meat, poultry, or fish you ate? *(Please do not include deep frying.)*

- ☐ NEVER (GO TO QUESTION 91)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

90a. Which of the following **fats** were regularly used to prepare your meat, poultry, or fish? ***(Mark all that apply.)***

- ☐ Margarine (including low-fat)
- ☐ Butter (including low-fat)
- ☐ Lard, fatback, or bacon fat
- ☐ Olive oil
- ☐ Corn oil
- ☐ Canola or rapeseed oil
- ☐ Oil spray, such as Pam or others
- ☐ Other kinds of oils
- ☐ None of the above

91. How often did you eat **tofu, soy burgers,** or **soy meat-substitutes**?

- ☐ NEVER (GO TO QUESTION 92)
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

91a. Each time you ate **tofu, soy burgers,** or **soy meat-substitutes**, how much did you usually eat?

- ☐ Less than ¼ cup or less than 2 ounces
- ☐ ¼ to ½ cup or 2 to 4 ounces
- ☐ More than ½ cup or more than 4 ounces

Question 92 appears in the next column

92. Over the <u>past 12 months</u>, did you eat **soups**?

- ☐ NO (GO TO QUESTION 93)
- ☐ YES

92a. How often did you eat **soup DURING THE WINTER**?

- ☐ NEVER
- ☐ 1–6 times per winter
- ☐ 7–11 times per winter
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

92b. How often did you eat **soup DURING THE REST OF THE YEAR**?

- ☐ NEVER
- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

92c. Each time you ate **soup**, how much did you usually eat?

- ☐ Less than 1 cup
- ☐ 1 to 2 cups
- ☐ More than 2 cups

92d. How often were the soups you ate **bean soups**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

92e. How often were the soups you ate **cream soups** (including chowders)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 93 appears on the next page

Over the <u>past 12 months</u>…

92f. How often were the soups you ate **tomato** or **vegetable soups**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

92g. How often were the soups you ate **broth soups** (including chicken) **with** or **without noodles** or **rice**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

93. How often did you eat **pizza**?

- ☐ NEVER (GO TO QUESTION 94)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

93a. Each time you ate **pizza**, how much did you usually eat?

- ☐ Less than 1 slice or less than 1 mini pizza
- ☐ 1 to 3 slices or 1 mini pizza
- ☐ More than 3 slices or more than 1 mini pizza

93b. How often did you eat pizza with **pepperoni, sausage**, or **other meat**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

94. How often did you eat **crackers**?

- ☐ NEVER (GO TO QUESTION 95)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 95 appears in the next column

94a. Each time you ate **crackers**, how many did you usually eat?

- ☐ Fewer than 4 crackers
- ☐ 4 to 10 crackers
- ☐ More than 10 crackers

95. How often did you eat **corn bread** or **corn muffins**?

- ☐ NEVER (GO TO QUESTION 96)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

95a. Each time you ate **corn bread** or **corn muffins,** how much did you usually eat?

- ☐ Less than 1 piece or muffin
- ☐ 1 to 2 pieces or muffins
- ☐ More than 2 pieces or muffins

96. How often did you eat **biscuits**?

- ☐ NEVER (GO TO QUESTION 97)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

96a. Each time you ate **biscuits**, how many did you usually eat?

- ☐ Fewer than 1 biscuit
- ☐ 1 to 2 biscuits
- ☐ More than 2 biscuits

97. How often did you eat **potato chips, tortilla chips,** or **corn chips** (including low-fat, fat-free, or low-salt)?

- ☐ NEVER (GO TO QUESTION 98)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 98 appears on the next page

Over the <u>past 12 months</u>…

97a. Each time you ate **potato chips, tortilla chips,** or **corn chips**, how much did you usually eat?

- ☐ Fewer than 10 chips or less than 1 cup
- ☐ 10 to 25 chips or 1 to 2 cups
- ☐ More than 25 chips or more than 2 cups

97b. How often were the chips you ate **Wow chips** or other **chips made with fat substitute** (Olean or Olestra)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

97c. How often were the chips you ate other **low-fat** or **fat-free chips**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

98. How often did you eat **popcorn** (including low-fat)?

- ☐ NEVER (GO TO QUESTION 99)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

98a. Each time you ate **popcorn**, how much did you usually eat?

- ☐ Less than 2 cups, popped
- ☐ 2 to 5 cups, popped
- ☐ More than 5 cups, popped

99. How often did you eat **pretzels**?

- ☐ NEVER (GO TO QUESTION 100)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 100 appears in the next column

99a. Each time you ate **pretzels**, how many did you usually eat?

- ☐ Fewer than 5 average twists
- ☐ 5 to 20 average twists
- ☐ More than 20 average twists

100. How often did you eat **peanuts, walnuts, seeds,** or **other nuts**?

- ☐ NEVER (GO TO QUESTION 101)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

100a. Each time you ate **peanuts, walnuts, seeds,** or **other nuts**, how much did you usually eat?

- ☐ Less than ¼ cup
- ☐ ¼ to ½ cup
- ☐ More than ½ cup

101. How often did you eat **energy, high-protein**, or **breakfast bars** such as **Power Bars, Balance, Clif,** or **others**?

- ☐ NEVER (GO TO QUESTION 102)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

101a. Each time you ate **energy, high-protein**, or **breakfast bars**, how much did you usually eat?

- ☐ Less than 1 bar
- ☐ 1 bar
- ☐ More than 1 bar

102. How often did you eat **yogurt** (NOT including frozen yogurt)?

- ☐ NEVER (GO TO QUESTION 103)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 103 appears on the next page

Over the <u>past 12 months</u>…

102a. Each time you ate **yogur**t, how much did you usually eat?

- ☐ Less than ½ cup or less than 1 container
- ☐ ½ to 1 cup or 1 container
- ☐ More than 1 cup or more than 1 container

103. How often did you eat **cottage cheese** (including low-fat)?

- ☐ NEVER (GO TO QUESTION 104)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

103a. Each time you ate **cottage cheese**, how much did you usually eat?

- ☐ Less than ¼ cup
- ☐ ¼ to 1 cup
- ☐ More than 1 cup

104. How often did you eat **cheese** (including low-fat; including on cheeseburgers or in sandwiches or subs)?

- ☐ NEVER (GO TO QUESTION 105)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

104a. Each time you ate **cheese**, how much did you usually eat?

- ☐ Less than ½ ounce or less than 1 slice
- ☐ ½ to 1 /2 ounces or 1 slice
- ☐ More than 1 /2 ounces or more than 1 slice

104b. How often was the cheese you ate **light** or **low-fat cheese**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

104c. How often was the cheese you ate **fat-free cheese**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

105. How often did you eat **frozen yogurt, sorbet,** or **ices** (including low-fat or fat-free)?

- ☐ NEVER (GO TO QUESTION 106)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

105a. Each time you ate **frozen yogurt, sorbet,** or **ices**, how much did you usually eat?

- ☐ Less than ½ cup or less than 1 scoop
- ☐ ½ to 1 cup or 1 to 2 scoops
- ☐ More than 1 cup or more than 2 scoops

106. How often did you eat **ice cream, ice cream bars,** or **sherbet** (including low-fat or fat-free)?

- ☐ NEVER (GO TO QUESTION 107)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

106a. Each time you ate **ice cream, ice cream bars,** or **sherbet**, how much did you usually eat?

- ☐ Less than ½ cup or less than 1 scoop
- ☐ ½ to 1 /2 cups or 1 to 2 scoops
- ☐ More than 1 /2 cups or more than 2 scoops

106b. How often was the ice cream you ate **light, low-fat,** or **fat-free ice cream** or **sherbet**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 105 appears in the next column

Question 107 appears on the next page

Over the <u>past 12 months</u>…

107. How often did you eat **cake** (including low-fat or fat-free)?

☐ NEVER (GO TO QUESTION 108)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

107a. Each time you ate **cake**, how much did you usually eat?

☐ Less than 1 medium piece
☐ 1 medium piece
☐ More than 1 medium piece

107b. How often was the cake you ate **light, low-fat,** or **fat-free cake**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

108. How often did you eat **cookies** or **brownies** (including low-fat or fat-free)?

☐ NEVER (GO TO QUESTION 109)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

108a. Each time you ate **cookies** or **brownies**, how much did you usually eat?

☐ Less than 2 cookies or 1 small brownie
☐ 2 to 4 cookies or 1 medium brownie
☐ More than 4 cookies or 1 large brownie

108b. How often were the cookies or brownies you ate **light, low-fat,** or **fat-free cookies** or **brownies**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 109 appears in the next column

109. How often did you eat **doughnuts, sweet rolls, Danish,** or **pop-tarts**?

☐ NEVER (GO TO QUESTION 110)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

109a. Each time you ate **doughnuts, sweet rolls, Danish,** or **pop-tarts**, how much did you usually eat?

☐ Less than 1 piece
☐ 1 to 2 pieces
☐ More than 2 pieces

110. How often did you eat **sweet muffins** or **dessert breads** (including low-fat or fat-free)?

☐ NEVER (GO TO QUESTION 111)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

110a. Each time you ate **sweet muffins or dessert breads**, how much did you usually eat?

☐ Less than 1 medium piece
☐ 1 medium piece
☐ More than 1 medium piece

110b. How often were the sweet muffins or dessert breads you ate **light, low-fat,** or **fat-free sweet muffins** or **dessert breads**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

111. How often did you eat **fruit crisp, cobbler,** or **strudel**?

☐ NEVER (GO TO QUESTION 112)

☐ 1–6 times per year	☐ 2 times per week
☐ 7–11 times per year	☐ 3–4 times per week
☐ 1 time per month	☐ 5–6 times per week
☐ 2–3 times per month	☐ 1 time per day
☐ 1 time per week	☐ 2 or more times per day

Question 112 appears on the next page

Over the <u>past 12 months</u>…

111a. Each time you ate **fruit crisp, cobbler,** or **strudel,** how much did you usually eat?

- ☐ Less than ½ cup
- ☐ ½ to 1 cup
- ☐ More than 1 cup

112. How often did you eat **pie**?

- ☐ NEVER (GO TO QUESTION 113)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

112a. Each time you ate **pie**, how much did you usually eat?

- ☐ Less than ⅛ of a pie
- ☐ About ⅛ of a pie
- ☐ More than ⅛ of a pie

The next four questions ask about the kinds of pie you ate. Please read all four questions before answering.

112b. How often were the pies you ate **fruit pie** (such as apple, blueberry, others)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

112c. How often were the pies you ate **cream, pudding, custard,** or **meringue pie**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

112d. How often were the pies you ate **pumpkin** or **sweet potato pie**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 113 appears in the next column

112e. How often were the pies you ate **pecan pie**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

113. How often did you eat **chocolate candy**?

- ☐ NEVER (GO TO QUESTION 114)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

113a. Each time you ate **chocolate candy**, how much did you usually eat?

- ☐ Less than 1 average bar or less than 1 ounce
- ☐ 1 average bar or 1 to 2 ounces
- ☐ More than 1 average bar or more than 2 ounces

114. How often did you eat **other candy**?

- ☐ NEVER (GO TO QUESTION 115)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

114a. Each time you ate **other candy**, how much did you usually eat?

- ☐ Fewer than 2 pieces
- ☐ 2 to 9 pieces
- ☐ More than 9 pieces

115. How often did you eat **eggs, egg whites,** or **egg substitutes** (NOT counting eggs in baked goods and desserts)? *(Please include eggs in salads, quiche, and soufflés.)*

- ☐ NEVER (GO TO QUESTION 116)

- ☐ 1–6 times per year
- ☐ 7–11 times per year
- ☐ 1 time per month
- ☐ 2–3 times per month
- ☐ 1 time per week
- ☐ 2 times per week
- ☐ 3–4 times per week
- ☐ 5–6 times per week
- ☐ 1 time per day
- ☐ 2 or more times per day

Question 116 appears on the next page

Over the <u>past 12 months</u>…

115a. Each time you ate **eggs**, how many did you usually eat?

- ☐ 1 egg
- ☐ 2 eggs
- ☐ 3 or more eggs

115b. How often were the eggs you ate **egg substitutes**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

115c. How often were the eggs you ate **egg whites only**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

115d. How often were the eggs you ate **regular whole eggs**?

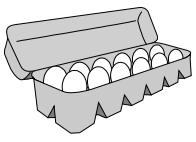
- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

115e. How often were the eggs you ate **cooked in oil, butter,** or **margarine**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

115f. How often were the eggs you ate part of **egg salad**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always



Question 116 appears in the next column

116. How many cups of **coffee**, caffeinated or decaffeinated, did you drink?

- ☐ NEVER (GO TO QUESTION 117)
- ☐ Less than 1 cup per month
- ☐ 1–3 cups per month
- ☐ 1 cup per week
- ☐ 2–4 cups per week
- ☐ 5–6 cups per week
- ☐ 1 cup per day
- ☐ 2–3 cups per day
- ☐ 4–5 cups per day
- ☐ 6 or more cups per day

116a. How often was the coffee you drank **decaffeinated**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

117. How many glasses of **ICED tea,** caffeinated or decaffeinated, did you drink?

- ☐ NEVER (GO TO QUESTION 118)
- ☐ Less than 1 cup per month
- ☐ 1–3 cups per month
- ☐ 1 cup per week
- ☐ 2–4 cups per week
- ☐ 5–6 cups per week
- ☐ 1 cup per day
- ☐ 2–3 cups per day
- ☐ 4–5 cups per day
- ☐ 6 or more cups per day

117a. How often was the iced tea you drank **decaffeinated** or **herbal tea**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

118. How many cups of **HOT tea**, caffeinated or decaffeinated, did you drink?

- ☐ NEVER (GO TO QUESTION 119)
- ☐ Less than 1 cup per month
- ☐ 1–3 cups per month
- ☐ 1 cup per week
- ☐ 2–4 cups per week
- ☐ 5–6 cups per week
- ☐ 1 cup per day
- ☐ 2–3 cups per day
- ☐ 4–5 cups per day
- ☐ 6 or more cups per day

118a. How often was the hot tea you drank **decaffeinated** or **herbal tea**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

Question 119 appears on the next page

Over the <u>past 12 months</u>…

119. How often did you add **sugar** or **honey** to your coffee or tea?

☐ NEVER (GO TO QUESTION 120)

☐ Less than 1 cup per month
☐ 1–3 cups per month
☐ 1 cup per week
☐ 2–4 cups per week
☐ 5–6 cups per week
☐ 1 cup per day
☐ 2–3 cups per day
☐ 4–5 cups per day
☐ 6 or more cups per day

119a. Each time **sugar** or **honey** was added to your coffee or tea, how much was usually added?

☐ Less than 1 teaspoon
☐ 1 to 3 teaspoons
☐ More than 3 teaspoons

120. How often did you add **artificial sweetener** to your coffee or tea?

☐ NEVER (GO TO QUESTION 121)

☐ Less than 1 time per month
☐ 1–3 times per month
☐ 1 time per week
☐ 2–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

120a. What kind of **artificial sweetener** did you usually use?

☐ Equal or aspartame
☐ Sweet N Low or saccharin

121. How often was **non-dairy creamer** added to your coffee or tea?

☐ NEVER (GO TO QUESTION 122)

☐ Less than 1 time per month
☐ 1–3 times per month
☐ 1 time per week
☐ 2–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

121a. Each time **non-dairy creamer** was added to your coffee or tea, how much was usually used?

☐ Less than 1 teaspoon
☐ 1 to 3 teaspoons
☐ More than 3 teaspoons

Question 122 appears in the next column

121b. What kind of **non-dairy creamer** did you usually use?

☐ Regular powdered
☐ Low-fat or fat-free powdered
☐ Regular liquid
☐ Low-fat or fat-free liquid

122. How often was **cream** or **half and half** added to your coffee or tea?

☐ NEVER (GO TO QUESTION 123)

☐ Less than 1 time per month
☐ 1–3 times per month
☐ 1 time per week
☐ 2–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

122a. Each time **cream** or **half and half** was added to your coffee or tea, how much was usually added?

☐ Less than 1 tablespoon
☐ 1 to 2 tablespoons
☐ More than 2 tablespoons

123. How often was **milk** added to your coffee or tea?

☐ NEVER (GO TO QUESTION 124)

☐ Less than 1 time per month
☐ 1–3 times per month
☐ 1 time per week
☐ 2–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2–3 times per day
☐ 4–5 times per day
☐ 6 or more times per day

123a. Each time **milk** was added to your coffee or tea, how much was usually added?

☐ Less than 1 tablespoon
☐ 1 to 3 tablespoons
☐ More than 3 tablespoons

123b. What kind of **milk** was usually added to your coffee or tea?

☐ Whole milk
☐ 2% milk
☐ 1% milk
☐ Skim, nonfat, or ½% milk
☐ Evaporated or condensed (canned) milk
☐ Soy milk
☐ Rice milk
☐ Other

Question 124 appears on the next page

Over the <u>past 12 months</u>…

124. How often was **sugar** or **honey** added to foods you ate? (Please do not include sugar in coffee, tea, other beverages, or baked goods.)

☐ NEVER (GO TO INTRODUCTION TO QUESTION 125)

☐ 1–6 times per year
☐ 7–11 times per year
☐ 1 time per month
☐ 2–3 times per month
☐ 1 time per week
☐ 2 times per week
☐ 3–4 times per week
☐ 5–6 times per week
☐ 1 time per day
☐ 2 or more times per day

124a. Each time **sugar** or **honey** was added to foods you ate, how much was usually added?

☐ Less than 1 teaspoon
☐ 1 to 3 teaspoons
☐ More than 3 teaspoons

The following questions are about the kinds of margarine, mayonnaise, sour cream, cream cheese, and salad dressing that you eat. If possible, please check the labels of these foods to help you answer.

125. Over the <u>past 12 months</u>, did you eat **margarine**?

☐ NO (GO TO QUESTION 126)

☐ YES

125a. How often was the margarine you ate **regular-fat margarine** (stick or tub)?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

125b. How often was the margarine you ate **light** or **low-fat margarine** (stick or tub)?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 126 appears in the next column

125c. How often was the margarine you ate **fat-free margarine**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

126. Over the past 12 months, did you eat **butter**?

☐ NO (GO TO QUESTION 127)

☐ YES

126a. How often was the butter you ate **light** or **low-fat butter**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

127. Over the <u>past 12 months</u>, did you eat **mayonnaise** or **mayonnaise-type dressing**?

☐ NO (GO TO QUESTION 128)

☐ YES

127a. How often was the mayonnaise you ate **regular-fat mayonnaise**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

127b. How often was the mayonnaise you ate **light** or **low-fat mayonnaise**?

☐ Almost never or never
☐ About ¼ of the time
☐ About ½ of the time
☐ About ¾ of the time
☐ Almost always or always

Question 128 appears on the next page

Over the past 12 months…

127c. How often was the mayonnaise you ate **fat-free mayonnaise**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

128. Over the past 12 months, did you eat **sour cream**?

- ☐ NO (GO TO QUESTION 129)
- ☐ YES

128a. How often was the sour cream you ate **regular-fat sour cream**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

128b. How often was the sour cream you ate **light, low-fat,** or **fat-free sour cream**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

129. Over the past 12 months, did you eat **cream cheese**?

- ☐ NO (GO TO QUESTION 130)
- ☐ YES

129a. How often was the cream cheese you ate **regular-fat cream cheese**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

129b. How often was the cream cheese you ate **light, low-fat,** or **fat-free cream cheese**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

130. Over the past 12 months, did you eat **salad dressing**?

- ☐ NO (GO TO INTRODUCTION TO QUESTION 131)
- ☐ YES

130a. How often was the salad dressing you ate **regular-fat salad dressing** (including oil and vinegar dressing)?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

130b. How often was the salad dressing you ate **light** or **low-fat salad dressing**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

130c. How often was the salad dressing you ate **fat-free salad dressing**?

- ☐ Almost never or never
- ☐ About ¼ of the time
- ☐ About ½ of the time
- ☐ About ¾ of the time
- ☐ Almost always or always

The following two questions ask you to summarize your usual intake of vegetables and fruits. Please do not include salads, potatoes, or juices.

131. Over the past 12 months, how many servings of **vegetables** (not including salad or potatoes) did you eat per week or per day?

- ☐ Less than 1 per week
- ☐ 1–2 per week
- ☐ 3–4 per week
- ☐ 5–6 per week
- ☐ 1 per day
- ☐ 2 per day
- ☐ 3 per day
- ☐ 4 per day
- ☐ 5 or more per day

Question 130 appears in the next column

Over the past 12 months…

132. Over the past 12 months, how many servings of **fruit** (not including juices) did you eat per week or per day?

- ☐ Less than 1 per week
- ☐ 1–2 per week
- ☐ 3–4 per week
- ☐ 5–6 per week
- ☐ 1 per day
- ☐ 2 per day
- ☐ 3 per day
- ☐ 4 per day
- ☐ 5 or more per day

133. Over the past month, which of the following foods did you eat **AT LEAST THREE TIMES?** ***(Mark all that apply.)***

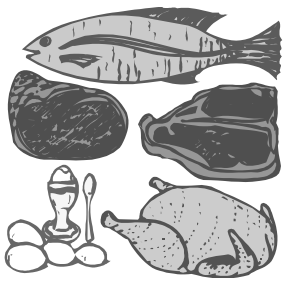
- ☐ Avocado, guacamole
- ☐ Cheesecake
- ☐ Chocolate, fudge, or butterscotch toppings or syrups
- ☐ Chow mein noodles
- ☐ Croissants
- ☐ Dried apricots
- ☐ Egg rolls
- ☐ Granola bars
- ☐ Hot peppers
- ☐ Jello, gelatin
- ☐ Milkshakes or ice-cream sodas
- ☐ Olives
- ☐ Oysters
- ☐ Pickles or pickled vegetables or fruit
- ☐ Plantains
- ☐ Pork neckbones, hock, head, feet
- ☐ Pudding or custard
- ☐ Veal, venison, lamb
- ☐ Whipped cream, regular
- ☐ Whipped cream, substitute
- ☐ NONE

134. For **ALL** of the past 12 months, have you followed any type of **vegetarian diet**?

- ☐ NO (GO TO INTRODUCTION TO QUESTION 135)
- ☐ YES

134a. Which of the following foods did you **TOTALLY EXCLUDE** from your diet? ***(Mark all that apply.)***

- ☐ Meat (beef, pork, lamb, etc.)
- ☐ Poultry (chicken, turkey, duck)
- ☐ Fish and seafood
- ☐ Eggs
- ☐ Dairy products (milk, cheese, etc.)



Introduction to Question 135 appears in the next column

The next questions are about your use of fiber supplements or vitamin pills.

135. Over the past 12 months, did you take any of the following types of **fiber or fiber supplements** on a regular basis (more than once per week for at least 6 of the last 12 months)? ***(Mark all that apply.)***

- ☐ NO, didn't take any fiber supplements on a regular basis (GO TO QUESTION 136)
- ☐ YES, psyllium products (such as Metamucil, Fiberall, Serutan, Perdiem, Correctol)
- ☐ YES, methylcellulose/cellulose products (such as Citrucel, Unifiber)
- ☐ YES, Fibercon
- ☐ YES, Bran (such as wheat bran, oat bran, or bran wafers)

136. Over the past 12 months, did you take any **multivitamins**, such as One-a-Day-, Theragran-, or Centrum-type multivitamins (as pills, liquids, or packets)?

- ☐ NO (GO TO INTRODUCTION TO QUESTION 138)
- ☐ YES

137. How often did you take **One-a-day-, Theragran-, or Centrum-type** multivitamins?

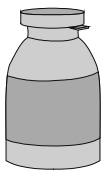
- ☐ Less than 1 day per month
- ☐ 1–3 days per month
- ☐ 1–3 days per week
- ☐ 4–6 days per week
- ☐ Every day

137a. Does your **multivitamin** usually contain **minerals** (such as iron, zinc, etc.)?

- ☐ NO
- ☐ YES
- ☐ Don't know

137b. For how many years have you taken **multivitamins**?

- ☐ Less than 1 year
- ☐ 1–4 years
- ☐ 5–9 years
- ☐ 10 or more years



Introduction to Question 138 appears on the next page

Over the <u>past 12 months</u>…

137c. Over the <u>past 12 months</u>, did you take any vitamins, minerals, or other herbal supplements other than your multivitamin?

☐ NO

Thank you *<u>very much</u>* for completing this questionnaire! Because we want to be able to use all the information you have provided, we would greatly appreciate it if you would please take a moment to review each page making sure that you:

- **Did not skip any pages and**
- **Crossed out the incorrect answer and circled the correct answer if you made any changes.**

☐ YES (GO TO INTRODUCTION TO QUESTION 138)

These last questions are about the vitamins, minerals, or herbal supplements you took that are <u>NOT</u> part of a One-a-day-, Theragran-, or Centrum-type of multivitamin.

Please include vitamins taken as part of an antioxidant supplement.

138. How often did you take **Beta-carotene** (**NOT** as part of a multivitamin in Question 137)?

☐ NEVER (GO TO QUESTION 139)

☐ Less than 1 day per month
☐ 1–3 days per month
☐ 1–3 days per week
☐ 4–6 days per week
☐ Every day

138a. When you took **Beta-carotene**, about how much did you take in one day?

☐ Less than 10,000 IU
☐ 10,000–14,999 IU
☐ 15,000–19,999 IU
☐ 20,000–24,999 IU
☐ 25,000 IU or more
☐ Don't know

138b. For how many years have you taken **Beta-carotene**?

☐ Less than 1 year
☐ 1–4 years
☐ 5–9 years
☐ 10 or more years

Question 139 appears in the next column

139. How often did you take **Vitamin A** (**NOT** as part of a multivitamin in Question 137)?

☐ NEVER (GO TO QUESTION 140)

☐ Less than 1 day per month
☐ 1–3 days per month
☐ 1–3 days per week
☐ 4–6 days per week
☐ Every day

139a. When you took **Vitamin A**, about how much did you take in one day?

☐ Less than 8,000 IU
☐ 8,000–9,999 IU
☐ 10,000–14,999 IU
☐ 15,000–24,999 IU
☐ 25,000 IU or more
☐ Don't know

139b. For how many years have you taken **Vitamin A**?

☐ Less than 1 year
☐ 1–4 years
☐ 5–9 years
☐ 10 or more years

140. How often did you take **Vitamin C** (**NOT** as part of a multivitamin in Question 137)?

☐ NEVER (GO TO QUESTION 141)

☐ Less than 1 day per month
☐ 1–3 days per month
☐ 1–3 days per week
☐ 4–6 days per week
☐ Every day

140a. When you took **Vitamin C**, about how much did you take in one day?

☐ Less than 500 mg
☐ 500–999 mg
☐ 1,000–1,499 mg
☐ 1,500–1,999 mg
☐ 2,000 mg or more
☐ Don't know

140b. For how many years have you taken **Vitamin C**?

☐ Less than 1 year
☐ 1–4 years
☐ 5–9 years
☐ 10 or more years

Question 141 appears on the next page

Over the <u>past 12 months</u>…

141. How often did you take **Vitamin E** (**NOT** as part of a multivitamin in Question 137)?

- ☐ NEVER (GO TO QUESTION 142)
- ☐ Less than 1 day per month
- ☐ 1–3 days per month
- ☐ 1–3 days per week
- ☐ 4–6 days per week
- ☐ Every day

141a. When you took **Vitamin E**, about how much did you take in one day?

- ☐ Less than 400 IU
- ☐ 400–799 IU
- ☐ 800–999 IU
- ☐ 1,000 IU or more
- ☐ Don't know

141b. For how many years have you taken **Vitamin E**?

- ☐ Less than 1 year
- ☐ 1–4 years
- ☐ 5–9 years
- ☐ 10 or more years

142. How often did you take **Calcium** or **Calcium-containing antacids** (**NOT** as part of a multivitamin in Question 137)?

- ☐ NEVER (GO TO QUESTION 143)
- ☐ Less than 1 day per month
- ☐ 1–3 days per month
- ☐ 1–3 days per week
- ☐ 4–6 days per week
- ☐ Every day

142a. When you took **Calcium** or **Calcium-containing antacids**, about how much elemental calcium did you take in one day? *(If possible, please check the label for elemental calcium.)*

- ☐ Less than 500 mg
- ☐ 500–599 mg
- ☐ 600–999 mg
- ☐ 1,000 mg or more
- ☐ Don't know

Question 143 appears in the next column

142b. For how many years have you taken **Calcium** or **Calcium-containing antacids**?

- ☐ Less than 1 year
- ☐ 1–4 years
- ☐ 5–9 years
- ☐ 10 or more years

The last two questions ask you about other supplements you took <u>more than once per week</u>.

143. Please mark any of the following **single supplements** you took <u>more than once per week</u> (**NOT** as part of a multivitamin in Question 137):

- ☐ B-6
- ☐ B-complex
- ☐ Brewer's yeast
- ☐ Cod liver oil
- ☐ Coenzyme Q
- ☐ Fish oil (Omega-3 fatty acids)
- ☐ Folic acid/folate
- ☐ Glucosamine
- ☐ Hydroxytryptophan (HTP)
- ☐ Iron
- ☐ Niacin
- ☐ Selenium
- ☐ Zinc

144. Please mark any of the following **herbal** or **botanical supplements** you took <u>more than once per week</u>.

- ☐ Aloe Vera
- ☐ Astragalus
- ☐ Bilberry
- ☐ Cascara sagrada
- ☐ Cat's claw
- ☐ Cayenne
- ☐ Cranberry
- ☐ Dong Kuai (Tangkwei)
- ☐ Echinacea
- ☐ Evening primrose oil
- ☐ Feverfew
- ☐ Garlic
- ☐ Ginger
- ☐ Ginkgo biloba
- ☐ Ginseng (American or Asian)
- ☐ Goldenseal
- ☐ Grapeseed extract
- ☐ Kava, kava
- ☐ Milk thistle
- ☐ Saw palmetto
- ☐ Siberian ginseng
- ☐ St. John's wort
- ☐ Valerian
- ☐ Other

Thank you *<u>very much</u>*** **for completing this questionnaire! Because we want to be able to use all the information you have provided, we would greatly appreciate it if you would please take a moment to review each page making sure that you:**

- **Did not skip any pages and**
- **Crossed out the incorrect answer and circled the correct answer if you made any changes.**

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Available online at www.sciencedirect.com
ScienceDirect
Mutation Research 650 (2008) 55–62
Genetic Toxicology and Environmental Mutagenesis
www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

Quantification of DNA repair capacity in whole blood of patients with head and neck cancer and healthy donors by comet assay

Daniel T. Saha [a], Bruce J. Davidson [b], Antai Wang [a], Allison J. Pollock [a], Roy A. Orden [a], Radoslav Goldman [a,*]

[a] *Georgetown University, Department of Oncology, Washington, DC, United States*
[b] *Georgetown University Hospital, Department of Otolaryngology Head and Neck Surgery, Washington, DC, United States*

Received 26 July 2007; received in revised form 9 October 2007; accepted 10 October 2007
Available online 18 October 2007

Abstract

Comet assay has been used to estimate cancer risk by quantification of DNA damage and repair in response to mutagen challenge. Our goal was to adopt best practices for the alkaline comet assay to measure DNA repair capacity of white blood cells in whole blood of patients with squamous cell carcinoma of the head and neck (HNSCC). The results show that initial damage by 10 Gy of gamma radiation expressed as percent DNA in comet tail was higher in stimulated lymphocytes (61.1 ± 11.8) compared to whole blood (43.0 ± 12.1) but subsequent repair was similar with comet tail of approximately 20% at 15 min and 13% at 45 min after exposure. Exposure of whole blood embedded in agarose from 5 to 10 Gy gamma radiation was followed by an approximately 70% repair of the DNA damage within 45 min with a faster repair phase in the first 15 min. Variability of the measurement was lower within repeated measurements of the same person compared to measurement of different healthy individuals. The repair during first 15 min was slower ($p = 0.01$) in ex-/non-smokers ($41.0 \pm 2.1\%$) compared to smokers ($50.3 \pm 2.7\%$). This phase of repair was also slower ($p = 0.02$) in HNSCC patients ($36.8 \pm 2.1\%$) compared to controls matched on age and smoking ($46.4 \pm 3.0\%$). The results of this pilot study suggest that quantification of repair in whole blood following a gamma radiation challenge is feasible. Additional method optimization would be helpful to improve the assay for a large population screening.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Comet assay; Whole blood; Squamous cell carcinoma of the head and neck

1. Introduction

Genome integrity is maintained by an intricate network of DNA repair proteins [1,2]. Defects in this complex machinery are associated with familial predispositions to cancer and other diseases [3]. Increasing evidence links environmental exposures, subtle modification in DNA repair efficiency, and cancer risk [4]. Establishing this connection, however, has been a challenge due to the complexity of interactions that affect the repair pathways [5,6]. Given the differences in DNA repair of lower eukaryotes, model organisms, and humans, molecular epidemiology is expected to provide important supportive evidence that is difficult to establish in targeted mechanistic studies [5,7]. Phenotypic cell-based assays of repair capacity provide a quantitative view of the complex pathways involved in repair [7,8]. Comet assay is one of the screens of

* Corresponding author at: Georgetown University, LCCC Room S183, 3970 Reservoir Road NW, Washington DC 20057, United States. Tel.: +1 202 687 9868; fax: +1 202 6871988.
E-mail address: rg26@georgetown.edu (R. Goldman).

1383-5718/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrgentox.2007.10.004

Author's personal copy

DNA integrity and repair capacity that is increasingly used in epidemiological studies exploring the connection of repair deficiency with cancer risk [9,10]. Comet assay quantifies single strand breaks and other damage in terms of migration of the DNA out of the nucleus under alkaline electrophoretic conditions [11]. Specific guidelines were developed to standardize the experimental conditions between laboratories [12]. This allows comparable inter-laboratory testing but further methodological development is needed to facilitate application of this assay to large population studies. In this paper, we describe adoption of best practices for the alkaline comet assay in an initial attempt to screen DNA repair of white blood cells in whole blood. An application of the assay to a population of controls and a small set of patients with squamous cell carcinoma of the head and neck (HNSCC) is discussed.

2. Materials and methods

2.1. Sample collection and processing

Blood samples were collected at the Georgetown University Hospital, Department of Otolaryngology Head and Neck Surgery and at the Lombardi Comprehensive Cancer Center. Age and smoking matched controls were recruited at Georgetown University in collaboration with the National Lung Screening Trial/Lung Screening Study (NLST/LSS) associated study "Recruitment of Control Participants for a Study of Head and Neck Cancer". All participants signed informed consent and samples were collected in accordance with guidelines of the Georgetown University Institutional Review Board. Patients with newly diagnosed squamous cell carcinoma of the head and neck were recruited into the study prior to receiving treatment. Participants were interviewed about lifestyle, medical and diet history. Blood samples for reproducibility studies were obtained from healthy volunteers. Blood samples were drawn in green-topped Vacutainer tubes (BD BioSciences, Franklin Lakes, NJ) containing sodium heparin (10–25 IU/mL). Blood samples for the study were processed according to two protocols: (1) whole blood processing; and (2) short-term culture of isolated lymphocytes (see below). Whole blood used for the comet assay was stored overnight at 4 °C. Prior to embedding in agarose, the blood samples were diluted 1:10 in RPMI 1640.

2.2. Study participants

To study the effects of smoking, we analyzed blood samples of a total of 40 non-cancer controls consisting of 17 smokers (mean age 61.8 ± 4.7 years) and 23 non- and ex-smokers (mean age 63.7 ± 5.9 years). To study the effects of head and neck cancer, we analyzed 12 cases (mean age 55.3 ± 12.4 years; 3 smokers (25%), 9 non- and ex-smokers) and 15 controls (mean age 57.9 ± 4.9 years; 4 smokers (26%), 11 non- and ex-smoker). Age of the two groups is not statistically different ($p = 0.46$).

2.3. Cell culture and cell exposures

Jurkat T-cells were cultured at 37 °C in 5% CO_2 in RPMI 1640 with 10% FBS and 1% sodium pyruvate. To obtain primary cultures, peripheral blood lymphocytes were isolated by gradient centrifugation using Histopaque-1077 (Sigma–Aldrich, St. Louis, MO) in Accuspin™ tubes (Sigma–Aldrich, St. Louis, MO). Isolated lymphocytes were seeded at a concentration of 5×10^5 cells/mL in RPMI 1640 (GIBCO-Invitrogen, Gaithersburg, MD) containing the following: 15% (v/v) fetal calf serum, 2 mM L-glutamine (GIBCO-Invitrogen, Gaithersburg, MD), 10 U/mL sodium heparin (Sigma–Aldrich, St. Louis, MO), 16 U/mL phytohemaglutinin (GIBCO-Invitrogen, Gaithersburg, MD), 1 U/mL penicillin and 1 μg/mL streptomycin (GIBCO-Invitrogen, Gaithersburg, MD). The cells were incubated at 37 °C in 5% CO_2 for 65 h. The cultured lymphocytes were then resuspended in media without PHA and supplemented with IL2 (20 U/mL) for an additional 28 h of culture.

DNA damage was induced using either the radiomimetic bleomycin or gamma radiation. For the bleomycin experiments, cells embedded in agarose were incubated for 30 min at 37 °C in RPMI 1640 containing 3 μg/mL bleomycin sulfate (Sigma–Aldrich, St. Louis, MO). For the irradiation experiments, cells were exposed to gamma rays derived from a Cs-137 source in a Research Irradiator. Cells in suspension in PBS were treated in a tissue culture flask (Corning Life Sciences, Corning, NY). Cells embedded in agarose on glass slides prior to irradiation were treated in a glass-staining dish (Thermo-Fisher, Waltham, MA) containing PBS (4 °C, pH 7.4; GIBCO-Invitrogen, Gaithersburg, MD).

2.4. Alkaline comet assay

The alkaline comet assay was conducted in accordance with the guidelines summarized by Tice et al. [12] with some modifications. Our comet slides were made in-house using a three-layer procedure where the cells were embedded in the topmost layer. Frosted slides (Thermo-Fisher, Waltham, MA) were coated in 1% agarose (Sigma–Aldrich, St. Louis, MO) and dried in an oven at 60 °C for 20 min. The second layer consisted of 0.75% normal melting agarose (Invitrogen, Gaithersburg, MD). Cells were mixed with 0.75% low-melting agarose (BioWhittaker-Cambrex, East Rutherford, NJ) kept at 37 °C under a yellow-frosted incandescent light (Philips Electronics North America, New York, NY) to prevent DNA damage. Both the second and third layers were flattened using a 22 mm × 22 mm coverslip (Thermo-Fisher, Waltham, MA) and the slides were placed on ice to facilitate the gelling process. Duplicate slides were made for each dose/repair point. After the embedding procedure, the cells were exposed to a mutagen to generate DNA damage. Following mutagen challenge, cells were either placed in lysis buffer (2.5 M NaCl, 100 mM EDTA, 10 mM Tris, 200 mM NaOH, 1% (v/v) Triton

Author's personal copy

X-100, 10% (v/v) DMSO) adjusted to pH 10 or allowed to repair. Cells allowed to repair the DNA damage were incubated in RPMI 1640 at 37 °C for a fixed length of time (typically 15 and 45 min) and then transferred to the lysis buffer, pH 10. The slides were kept in the lysis buffer at 4 °C in the dark for 3–24 h. The slides were transferred to ice-cold electrophoresis buffer (pH 13; 300 mM NaOH, 1 mM EDTA) for 40 min. Electrophoresis was carried out at 1.3 V/cm for 45 min at 4 °C in the dark. The volume of the buffer was adjusted so that the current at the start was 300 mA. Slides were neutralized using three 5-min washes with autoclaved 0.4 M Tris buffer (pH 7.4). DNA was fixed with methanol for 10 min. Slides were then washed with autoclaved distilled water twice for 5 min and allowed to dry overnight. The dried slides kept in a storage box (Thermo-Fisher, Waltham, MA) until scoring.

2.5. Comet scoring and data collection

All samples were coded to blind the experimenter to case/control status. Additionally, the dose and repair information on the slides was masked prior to scoring to prevent bias in the selection of comets for imaging. Slides were rehydrated for 45–60 min in autoclaved distilled water and stained for 10 min using a 1 μg/mL ethidium bromide solution (Invitrogen, Gaithersburg, MD). Excess stain was washed away using three 5 min washes with autoclaved distilled water. Slides were scored wet with a cover glass over the gel. Comet images were obtained using an Olympus BX-51 microscope (Opelco-Olympus, Center Valley, PA) equipped with a 100 W mercury burner and a wide green fluorescent mirror (U-MWG2) and a cooled 5 megapixel digital CCD camera (QImaging Micropublisher 5.0 RTV, QImaging, Surrey, BC, Canada). Two neutral density filters (U-25ND25 and U-25ND50) were inserted into the light path to reduce photobleaching of the fluorophore. Images were scored using a semi-automated Comet Analysis System (Loats Associates, Westminster, MD). A minimum of 50 images were recorded at 400× total magnification for each slide with two slides per dose–repair point. In accordance with the idea that "hedgehogs" were dead cells that do not offer information regarding DNA repair [13], we tracked them as a categorical variable but did not include them in the calculation of the summary statistics. We observed 2.1 "hedgehogs" per 100 cells on average. The Loats software calculated a number of parameters but we focused primarily on percent DNA in tail, the most reliable parameter for inter-laboratory comparisons [14]. For completeness, data on tail length and Olive tail moment is provided in Supplementary materials.

2.6. Data analysis

One hundred comet images were recorded for each dose–repair point (2 slides 50 images each). These images were analyzed by a semi automated scoring system (Loats Associates, Westminster, MD). The percent DNA in tail for each image was used as the variable of interest. Experiments involving cell lines were typically repeated a minimum of three times and the summary statistics (mean, standard deviation, median, range, and variance) were calculated according to standard formulas. Experiments using whole blood samples from the population were done once per sample. The sampling distribution of the mean and standard error for this population of samples was computed for each dose–repair point. Percent repaired was calculated as follows: [(Initial damage − Damage at time T)/Initial damage] × 100. ANOVA was used to check for differences between cell types (Jurkat T-cells, primary lymphocytes, and whole blood) and for differences between bleomycin-induced damage and gamma radiation-induced damage.

To determine the variability "within" a sample, the assay was conducted nine times on blood samples obtained from one control. To determine the variability between samples, 40 healthy controls were selected for analysis. The effects of smoking were examined by sorting the 40 controls by smoking status (current smoker *vs*. ex-smoker/non-smoker). Ex-smokers were defined as subjects that quit smoking for at least 6 months; non-smokers smoked less than 100 cigarettes during their life time. Data from a small pilot study of 27 men and women recruited for a study of HNSCC (12 HNSCC cases and 15 cancer free controls) was examined for differences in DNA repair capacity. A two-sided t-test was used to compare the mean "percent DNA in tail" measurements for each dose–repair point as well as the percent repaired at 15 and 45 min. The p-values reported in the text are based on two-sided t-test unless otherwise indicated.

3. Results

Jurkat T-cell line was used to establish the comet assay for comparison of lymphocyte and whole blood assays in our laboratory. We focused on DNA repair measured at several time points following a mutagen challenge. Comparison of repair in Jurkat T-cells, stimulated primary lymphocytes, and whole blood is presented in Fig. 1. Jurkat T-cells showed the highest initial damage following exposure to 10 Gy gamma radiation with progressively less damage observed in primary lymphocytes and whole blood. Jurkat T-cells also showed fastest DNA repair with approximately 90% of the DNA damage repaired within 15 min. The repair in primary lymphocytes and whole blood was similar (no significant difference at 8–45 min after exposure) with both repairing approximately 70% damage within 45 min. As reported previously, the repair kinetic appears biphasic with a faster repair within first 15 min and a slower repair from 15 to 45 min [9,15]. Except for the Jurkat T, the cells did not completely repair the damage even at 45 min. To minimize variability of the repair measurement, we evaluated the possibility to embed the cells into agarose prior to exposure [16,17]. The results for lymphocytes suspended in PBS in tissue culture flasks

Author's personal copy

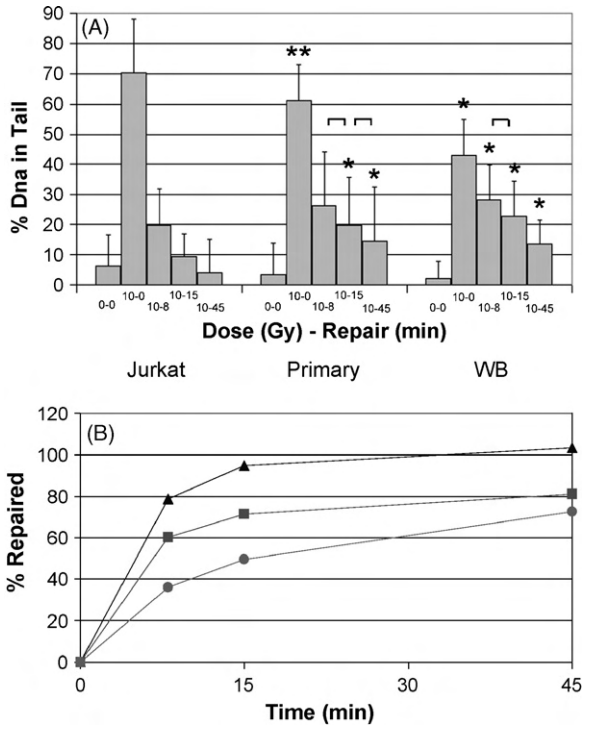


Fig. 1. Kinetic of repair over 45 min at 37 °C in RPMI media in three types of cell preparations ($n = 3$ for each) exposed to 10 Gy of gamma radiation. Whole blood was obtained from a healthy (non-cancer) volunteer. (A) *Significantly different from the corresponding dose-time point in Jurkat T-cells ($p < 0.05$). **Significantly different from the corresponding dose-time point in Jurkat T-cells and whole blood ($p < 0.05$). Brackets indicate the only differences within each series that were not significantly different. (B) Percent of repaired DNA in tail compared to 10 Gy 0 min. Triangles, Jurkat T-cells; squares, primary cultured lymphocytes; circles, whole blood.

or embedded in agarose prior to irradiation were nearly identical (Table 1). A uniform quick start and end of the repair process was achieved by transfer of the slides with embedded cells from PBS to RPMI at the start and from RPMI to lysis buffer at the end of the assay. We found similar results for Jurkat T-cells and whole blood (data not shown). These results suggest that the cells embedded in agarose can be used for quantification of repair. Because of similar repair kinetic in whole blood and cultured lymphocytes and because of ease of handling of whole blood, we chose to continue our experiments with embedded whole blood. For all experiments, whole blood was analyzed after an overnight storage at 4 °C [18,19] which was previously reported to minimize variability of the measurement [20].

Fig. 2 shows a dose response for whole blood embedded in agarose prior to treatment with gamma radiation. The initial damage increased linearly with dose ($r^2 = 0.92$). Our criteria for an "optimum dose"

Table 1
Comparison of repair following exposure of primary cultured lymphocytes to 10 Gy of gamma radiation

Dose (Gy)–repair (min)	Embedded Mean ± S.D.	Suspension Mean ± S.D.
0–0	3.40 ± 10.51	4.71 ± 12.54
0–8	1.97 ± 5.30	4.11 ± 12.81
0–15	2.00 ± 7.96	5.10 ± 14.59
0–45	2.95 ± 10.18	5.38 ± 13.12
10–0	61.16 ± 11.79	60.23 ± 14.07
10–8	26.32 ± 17.84	31.34 ± 21.64
10–15	19.9 ± 15.85	20.38 ± 17.62
10–45	14.28 ± 18.35	13.52 ± 15.69

Lymphocytes were either embedded in agarose or in suspension during irradiation. Cells treated in suspension were embedded in agarose prior to repair. Values shown are the mean ± S.D. for three replicates of each condition.

included initial damage close to 50% and a high ratio of initial damage-to-damage at 45 min. Both 9 and 10 Gy of gamma irradiation induced about 50% DNA in tail with 9 Gy showing the greatest differential at 45 min. The cells repaired about 70% of the initial damage at 45 min for almost every dose which suggests that majority of the cells are able to repair the damage induced by increasing doses of radiation. This supports the notion that the embedded cells are intact and functioning during the repair period.

In addition to ionizing radiation, the radiomimetic bleomycin is often used in studies of DNA damage and repair [9,21]. We compared treatment of whole blood with bleomycin, 3 μg/mL, to treatment with gamma radiation. Bleomycin, induced about 90% DNA in tail, compared to 40% DNA in tail after 10 Gy of gamma radiation (Fig. 3). The repair of the bleomycin-induced damage was fast, with about 70% of the damage repaired at 15 min. Bleomycin exposure would be therefore a

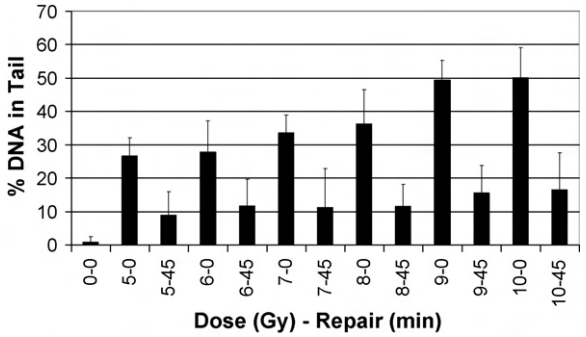


Fig. 2. Dose response of whole blood from a healthy (non-cancer) volunteer to gamma radiation exposure. Each bar represents average percent DNA in tail at a given dose (0–10 Gy) at two time points (0 or 45 min of repair).

Author's personal copy

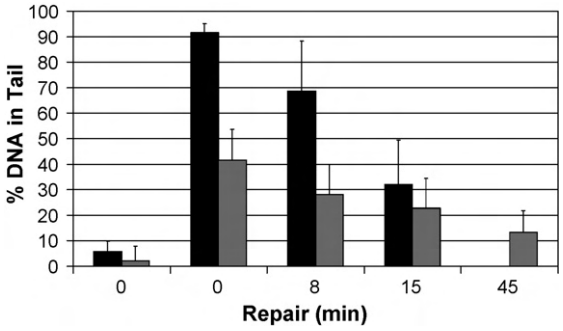


Fig. 3. Exposure of whole blood from a healthy (non-cancer) volunteer to bleomycin and gamma radiation. Black bars, cells exposed to bleomycin, 3 μg/mL; dotted bars, cells exposed to gamma irradiation, 10 Gy.

good model for evaluation of the repair kinetic but we observed a large day-to-day variability in the bleomycin experiments (data not shown). The dosing with gamma radiation was therefore selected for further experiments. The damage within each series (bleomycin and gamma radiation) was significantly different at all time points except the 8 and 15 min time points in the gamma radiation series. The 8 min time point was not included in further experiments. The day-to-day variability of the assay following gamma radiation was evaluated on whole blood from a single donor measured nine separate times (Table 2). The subject-to-subject variability was measured by performing the assay once for each of 40 non-cancer controls. The means and medians for each dose–repair point were similar between the two sets of data but the variance and range (maximum minus minimum) were smaller for each dose–repair point in the day-to-day measurements compared to measurements between subjects. The higher inter individual variability suggests that the assay could provide a useful comparison of repair in a population.

Table 2
Variability of the measurement determined by the following: (A) "within nine replicates", day-to-day variability of the whole blood assay measured nine times on a single healthy volunteer; and (B) "between 40 non-cancer controls", subject-to-subject variability of the whole blood assay measured once for each of 40 population controls

Dose (Gy)–repair (min)	9–0	9–15	9–45
(A) Within nine replicates			
Mean ± S.D.	41.95 ± 7.17	22.74 ± 4.59	10.53 ± 4.77
Variance	51.42	21.04	22.79
Median	40.53	23.11	8.78
Range	27.22	15.33	14.30
(B) Between 40 non-cancer controls			
Mean ± S.E.M.	44.17 ± 1.45	24.84 ± 1.14	15.76 ± 0.93
Variance	83.56	51.74	34.66
Median	44.06	25.32	16.02
Range	39.99	26.57	22.00

S.D., standard deviation; S.E.M., standard error of the mean; range, maximum minus minimum.

To evaluate the effect of smoking on DNA repair, forty non-cancer controls used in the measurement of variability were sorted according to smoking status. Non-smokers were grouped with ex-smokers and compared to current smokers. The values for percent DNA in tail for each dose–repair point were similar between the two groups (Table 3). The percent repair of DNA in tail between 9–0 and 9–15 min, i.e. the fast phase of the repair kinetic, was higher for smokers than non-/ex-smokers ($p = 0.01$). This observation suggests that single strand break repair might be induced in white blood cells by the exposure to cigarette smoke.

The optimized assay conditions were evaluated in a pilot study of HNSCC. We compared HNSCC cases ($n = 12$) with controls frequency matched on age and smoking ($n = 15$). HNSCC cases consisted of 3 smokers (25%) and 9 non-/ex-smokers with mean age 55.3

Table 3
Effect of smoking on the repair kinetic in forty non-cancer controls

Dose (Gy) -Repair (min)	9–0	9–15	9–45	% Repair 0–15	% Repair 15–45
Non-smokers (4) and ex-smokers (19)					
Mean	41.83	24.69	15.84	40.98	21.26
S.E.M.	1.94	1.37	1.14	2.13	1.82
Smokers (17)					
Mean	40.90	20.76	12.31	50.30	21.09
S.E.M.	3.11	1.98	1.34	2.74	2.18
t-Test					
p-value	0.27	0.40	0.18	0.01	0.95

Whole blood of 17 smokers and 23 non-/ex-smokers embedded in agarose was exposed to 9 Gy gamma radiation following overnight storage at 4 °C. 9–0, 9 Gy 0 min of repair; 9–15, 9 Gy 15 min of repair; 9–45, 9 Gy 45 min of repair; % repair for 0–15 and 15–45 min was calculated as described in Section 2. *p*-values are based on a two-sided *t*-test.

Author's personal copy

Table 4
Effect of HNSCC on the repair kinetic

Dose (Gy)–repair (min)	9–0	9–15	9–45	% Repair 0–15	% Repair 15–45
Cases ($n = 12$)					
Mean ± S.E.M.	46.88 ± 3.58	29.98 ± 2.71	19.04 ± 1.85	36.77 ± 2.14	22.26 ±± 3.34
Median	51.13	31.38	18.52	35.46	27.32
Range	34.20	31.57	23.14	24.79	37.55
Controls ($n = 15$)					
Mean ± S.E.M.	43.27 ± 2.70	23.02 ± 1.80	13.93 ± 1.71	46.37 ± 2.99	21.84 ± 2.84
Median	44.87	21.70	11.95	41.48	20.53
Range	39.37	26.06	21.77	46.54	48.61
t-Test					
p-values	0.43	0.04	0.06	0.02	0.93

Whole blood of 12 HNSCC patients and 15 cancer free controls embedded in agarose was exposed to 9 Gy gamma radiation following overnight storage at 4 °C. 9–0, 9 Gy 0 min of repair; 9–15, 9 Gy 15 min of repair; 9–45, 9 Gy 45 min of repair; % repair for 0–15 and 15–45 min was calculated as described in Section 2. *p*-values are based on a two-sided *t*-test.

years; controls consisted of 4 current smokers (26%) and 11 non-/ex-smokers with mean age 57.9 years. Age of the two groups is not statistically different ($p = 0.46$). Two participants (one case and one control) did not provide exact quit date to determine current or ex-smoker categories. The percent repaired at 15 min was significantly lower in HNSCC cases ($p = 0.02$) and percent DNA in the comet tail at 15 and 45 min of repair was marginally higher in HNSCC patients (Table 4). This small pilot supports the previously observed association of lower DNA repair capacity with increased risk of HNSCC [21–23]. The study would have to be expanded to confirm this pilot observation.

4. Discussion

Measurement of DNA strand breaks by comet assay has been used to evaluate genotoxicity, to monitor exposures, or to quantify DNA damage and repair in molecular epidemiology [24–26]. In particular, comet assay has been used to examine the association of DNA repair capacity in peripheral white blood cells with risk of various cancers [9,10,17] including HNSCC [23,27]. Palyvoda et al. [27] measured DNA damage at six time points between 0 and 180 min after gamma radiation exposure in 48 HNSCC patients and 38 healthy controls. The study showed increased baseline damage and decreased repair in HNSCC patients. The authors also reported high variability in background DNA damage which suggests that standardization of the assay conditions is important. Iwakawa et al. [23] compared residual DNA damage in 10 healthy controls and 87 HNSCC patients. The mean residual damage after 15 min of repair was significantly higher ($p < 0.01$) in HNSCC cases (56.8 + 24.4) compared to healthy controls (42.9 + 19.6). This study also showed that such differences were not observed in EBV-transformed cell lines. In this study, we present initial results of our attempt to adopt standard assay condition for quantification of DNA repair in HNSCC patients and controls.

The monitoring of repair in the population would be ideally done on target tissue. This is not practical and peripheral white blood cells are typically used as a surrogate. The assumption is that the DNA repair capacity of an individual is a genetic predisposition measurable in various cell types. This notion is supported by the results of studies of relatives and twins showing that the repair phenotypes are heritable [8,21]. Besides the use of isolated lymphocytes, several studies reported use of whole blood for the population studies [19,23]. Additional purification and culturing of the lymphocytes does not provide a clear advantage; it is not clear at present which subtype would be a closer approximation of the target tissue or why. We therefore adopted analysis of whole blood in our study. The performance of the assay on freshly isolated blood samples is complicated by the timing of patient sample collection and inability to store/repeat analyses. It would be advantageous to use cryo-preserved or EBV-transformed cells. Our experience and publications of others show, however, that the measurement of DNA repair in cryo-preserved and EBV-transformed cells is not feasible at present [23,28].

Bleomycin is a radiomimetic commonly used in DNA repair studies [9,21]. In our hands, the initial damage in bleomycin treated cells is high and repaired fast (Fig. 3). However, the day-to-day variability is higher than in the gamma radiation experiments which prompted us to select the exposure to radiation for the population study.

In our study, the highest initial damage and the fastest repair was observed in Jurkat T-cells. It appears that these

Author's personal copy

actively replicating cells are most sensitive to gamma radiation, followed by the lymphocytes stimulated with PHA and IL-2. The whole blood is less sensitive possibly due to higher fraction of the cells in G_0 [29]. The repair in primary cultured lymphocytes and fresh whole blood is comparable as demonstrated in Fig. 1. The repair was about 70% complete at 45 min at all doses examined (Fig. 2). We chose a dose of 9 Gy for the study of DNA repair capacity; this dose is comparable with other studies of DNA repair [14,17]. Initially, we tested multiple time points along the repair kinetic (data not shown) but for analysis of patient samples we chose two time points (15 and 45 min) to approximate the fast and slow portion of the repair [9,15]. Analysis of additional time points in a large population study would not, in our experience, be feasible. With two repair time points, we are able to process slides for a maximum of four patients at a time with a minimum of 2 days needed to complete the sample processing.

The workflow selected based on the above results includes sampling of whole blood with an overnight storage at 4 °C, exposure of cells embedded in agarose to 9 Gy of gamma radiation, and measurement of repair at 15 and 45 min after exposure. Under these conditions, the variability of the measurement between people is higher than the day-to-day variability for the same person sampled repeatedly (Table 2). The entire workflow focuses on quantification of repair as opposed to quantification of initial damage. The disappearance of the comet tail is only a rough approximation of the repair process and does not evaluate in detail any of the relevant enzymatic processes. In addition, the phenotype may be further influenced by factors only indirectly connected with DNA repair including chromatin structure [30]. However, such a global view may generate hypotheses stimulating further research of the repair pathways [7]. We used these conditions to compare DNA repair in smokers and in HNSCC patients. The repair was slightly higher during the first 15 min in smokers compared to ex-/non-smokers (Table 3). This is possibly due to induction of repair by exposure to cigarette smoke. The repair was also slightly higher in controls compared to HNSCC patients matched on age and smoking status (Table 4). This could reflect a genetic predisposition similar to the heritable phenotype of mutagen sensitivity previously associated with increased risk of HNSCC [22].

In conclusion, our results show that the described workflow is appropriate for quantification of DNA repair on blood samples in a population study. We were able to detect a slower repair in HNSCC patients in a small pilot study of DNA repair. However, the throughput of the assay is limited and the application to a large population would benefit from further methodological improvements including an efficient cell preservation protocol and automated image processing.

Acknowledgments

We want to thank Dr. Bozena Novotna for an expert introduction to comet assay analyses. The project was supported in part by grant number M01RR-023942-01 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH. Additional support was provided by U.S. Army Medical Research and Material Command, Prostate Cancer Research Program award W81XWH-04-1-0294, American Cancer Society award CRTG-02-245-01-CCE, and Flight Attendant Medical Research Institute award 052444 to R.G. Most importantly, we acknowledge the study participants for their contributions to making this study possible.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mrgentox.2007.10.004.

References

[1] T. Lindahl, R.D. Wood, Quality control by DNA repair, Science 286 (1999) 1897–1905.
[2] J.H. Hoeijmakers, Genome maintenance mechanisms for preventing cancer, Nature 411 (2001) 366–374.
[3] V.A. Bohr, DNA damage and its processing: relation to human disease, J. Inherit. Metab. Dis. 25 (2002) 215–222.
[4] H.W. Mohrenweiser, D.M. Wilson III, I.M. Jones, Challenges and complexities in estimating both the functional impact and the disease risk associated with the extensive genetic variation in human DNA repair genes, Mutat. Res. 526 (2003) 93–125.
[5] D.E. Barnes, T. Lindahl, Repair and genetic consequences of endogenous DNA base damage in mammalian cells, Annu. Rev. Genet. 38 (2004) 445–476.
[6] B. Vogelstein, K.W. Kinzler, Cancer genes and the pathways they control, Nat. Med. 10 (2004) 789–799.
[7] I.M. Jones, C.B. Thomas, T. Xi, H.W. Mohrenweiser, D.O. Nelson, Exploration of methods to identify polymorphisms associated with variation in DNA repair capacity phenotypes, Mutat. Res. 616 (2007) 213–220.
[8] M. Berwick, P. Vineis, Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review, J. Natl. Cancer Inst. 92 (2000) 874–897.
[9] P. Schmezer, N. Rajaee-Behbahani, A. Risch, S. Thiel, W. Rittgen, P. Drings, H. Dienemann, K.W. Kayser, V. Schulz, H. Bartsch, Rapid screening assay for mutagen sensitivity and DNA repair capacity in human peripheral blood lymphocytes, Mutagenesis 16 (2001) 25–30.

Author's personal copy

[10] M.B. Schabath, M.R. Spitz, H.B. Grossman, K. Zhang, C.P. Dinney, P.J. Zheng, X. Wu, Genetic instability in bladder cancer assessed by the comet assay, J. Natl. Cancer Inst. 95 (2003) 540–547.
[11] N.P. Singh, M.T. McCoy, R.R. Tice, E.L. Schneider, A simple technique for quantitation of low levels of DNA damage in individual cells, Exp. Cell Res. 175 (1988) 184–191.
[12] R.R. Tice, E. Agurell, D. Anderson, B. Burlinson, A. Hartmann, H. Kobayashi, Y. Miyamae, E. Rojas, J.C. Ryu, Y.F. Sasaki, Single cell gel/comet assay: guidelines for in vitro and in vivo genetic toxicology testing, Environ. Mol. Mutagen. 35 (2000) 206–221.
[13] B. Burlinson, R.R. Tice, G. Speit, E. Agurell, S.Y. Brendler-Schwaab, A.R. Collins, P. Escobar, M. Honma, T.S. Kumaravel, M. Nakajima, Y.F. Sasaki, V. Thybaud, Y. Uno, M. Vasquez, A. Hartmann, Fourth International Workgroup on Genotoxicity testing: Results of the in vivo Comet assay workgroup, Mutat. Res./Genet. Toxicol. Environ. Mutagen. 627 (2007) 31–35.
[14] T.S. Kumaravel, A.N. Jha, Reliable Comet assay measurements for detecting DNA damage induced by ionising radiation and chemicals, Mutat. Res. 605 (2006) 7–16.
[15] M. Frankenberg-Schwager, Review of repair kinetics for DNA damage induced in eukaryotic cells in vitro by ionizing radiation, Radiother. Oncol. 14 (1989) 307–320.
[16] J.P. McNamee, J.R. McLean, C.L. Ferrarotto, P.V. Bellier, Comet assay: rapid processing of multiple samples, Mutat. Res. 466 (2000) 63–69.
[17] T.R. Smith, M.S. Miller, K.K. Lohman, L.D. Case, J.J. Hu, DNA damage and breast cancer risk, Carcinogenesis 24 (2003) 883–889.
[18] D. Anderson, T.W. Yu, M.M. Dobrzynska, G. Ribas, R. Marcos, Effects in the comet assay of storage conditions on human blood, Teratogen. Carcinogen. Mutagen. 17 (1997) 115–125.
[19] C.H. Chuang, M.L. Hu, Use of whole blood directly for single-cell gel electrophoresis (comet) assay in vivo and white blood cells for in vitro assay, Mutat. Res. 564 (2004) 75–82.
[20] B. Novotna, Increased DNA fragmentation detected by comet assay in peripehral blood of heterozygotes for a frameshift mutation in the DNA repair protein nibrin. DNA Repair Workshop, Smolenice, Slovakia, 2000, abstract p. 33.
[21] X. Wu, J. Gu, M.R. Spitz, Mutagen sensitivity: a genetic predisposition factor for cancer, Cancer Res. 67 (2007) 3493–3495.
[22] J. Cloos, M.R. Spitz, S.P. Schantz, T.C. Hsu, Z.F. Zhang, H. Tobi, B.J. Braakhuis, G.B. Snow, Genetic susceptibility to head and neck squamous cell carcinoma, J. Natl. Cancer Inst. 88 (1996) 530–535.
[23] M. Iwakawa, M. Goto, S. Noda, M. Sagara, S. Yamada, N. Yamamoto, Y. Kawakami, Y. Matsui, Y. Miyazawa, H. Yamazaki, H. Tsuji, T. Ohno, J. Mizoe, H. Tsujii, T. Imai, DNA repair capacity measured by high throughput alkaline comet assays in EBV-transformed cell lines and peripheral blood cells from cancer patients and healthy volunteers, Mutat. Res. 588 (2005) 1–6.
[24] A.R. Collins, E. Horvathova, Oxidative DNA damage, antioxidants and DNA repair: applications of the comet assay, Biochem. Soc. Trans. 29 (2001) 337–341.
[25] E. Rojas, M.C. Lopez, M. Valverde, Single cell gel electrophoresis assay: methodology and applications, J. Chromatogr. B Biomed. Sci. Appl. 722 (1999) 225–254.
[26] P.L. Olive, J.P. Banath, The comet assay: a method to measure DNA damage in individual cells, Nat. Protocols 1 (2006) 23–29.
[27] O. Palyvoda, I. Mukalov, J. Polanska, A. Wygoda, L. Drobot, M. Widel, J. Rzeszowska-Wolny, Radiation-induced DNA damage and its repair in lymphocytes of patients with head and neck cancer and healthy donors, Anticancer Res. 22 (2002) 1721–1725.
[28] S.J. Duthie, L. Pirie, A.M. Jenkinson, S. Narayanan, Cryopreserved versus freshly isolated lymphocytes in human biomonitoring: endogenous and induced DNA damage, antioxidant status and repair capability, Mutagenesis 17 (2002) 211–214.
[29] S.W. Maluf, Monitoring DNA damage following radiation exposure using cytokinesis-block micronucleus method and alkaline single-cell gel electrophoresis, Clin. Chim. Acta 347 (2004) 15–24.
[30] A.J. Morrison, X. Shen, DNA repair in the context of chromatin, Cell Cycle 4 (2005) 568–571.

Carcinogenesis vol.28 no.10 pp.2149–2153, 2007
doi:10.1093/carcin/bgm177
Advance Access publication August 27, 2007

Candidate markers for the detection of hepatocellular carcinoma in low-molecular weight fraction of serum

Radoslav Goldman*, Habtom W.Ressom, Mohamed Abdel-Hamid[1], Lenka Goldman, Antai Wang, Rency S.Varghese, Yanming An, Christopher A.Loffredo, Steven K.Drake[2], Sohair A.Eissa[3], Iman Gouda[3], Sameera Ezzat[4] and Francoise Seillier Moiseiwitsch

Georgetown University, Lombardi Comprehensive Cancer Center, 3970 Reservoir Road Northwest, Washington, DC 20057, USA, [1]Minia University and Viral Hepatitis Research Laboratory, NHTMRI, Cairo, Egypt, [2]Clinical Chemistry Service, Department of Laboratory Medicine, National Institutes of Health, Bethesda, MD, USA, [3]National Cancer Institute, Cairo, Egypt and [4]Menoufiya University, Shibin El Kom, Egypt

*To whom correspondence should be addressed. Tel: +1 202 687 9868; Fax: +1 202 687 1988;
Email: rg26@georgetown.edu

Hepatocellular carcinoma (HCC) represents an important public health problem in Egypt where up to 90% of HCC cases are attributable to hepatitis C viral (HCV) infection. Serum alpha-fetoprotein is elevated in only ~60% of HCC patients. The development of effective markers for the detection of HCC could have an impact on cancer mortality and significant public health implications worldwide. The objective of our study was to assess six candidate markers for detection of HCC identified by mass spectrometric analysis of enriched serum. The study examined 78 HCC cases and 72 age- and gender-matched cancer-free controls recruited from the Egyptian population. Matrix-assisted laser desorption–ionization time-of-flight mass spectrometric analysis of enriched low-molecular weight fraction of serum was used for identification of the candidate markers. Our analyses show that all six candidate markers are associated with HCC after adjustment for important covariates including HCV and hepatitis B viral infections. The marker candidates are independently predictive of HCC with areas under the receiver operating characteristic (AuROC) curve ranging from 63–93%. A combination of the six markers improves prediction accuracy to 100% sensitivity, 91% specificity and 98% AuROC curve in an independent test set of 50 patients. Two of the candidate markers were identified by sequencing as fragments of complement C3 and C4. In conclusion, a set of six peptides distinguished with high prediction accuracy HCC from controls in an Egyptian population with a high rate of chronic HCV infection. Further evaluation of these marker candidates for the diagnosis of HCC is needed.

Introduction

Hepatocellular carcinoma (HCC) is a significant worldwide health problem with as many as 500 000 new cases diagnosed each year (1). There is considerable geographical variation in the incidence of HCC (2). In Egypt, HCC is third among cancers in men with >8000 new cases predicted by 2012 (3–5). The HCC epidemic in Egypt is associated with hepatitis C viral (HCV) infection; Egypt has the highest prevalence of HCV in the world with ~13.8% of the population infected and seven million with chronic HCV liver disease (6). Up to 90% of HCC cases in the Egyptian population were attributed to HCV (5,7). In the USA, the increasing incidence of HCC has been associated with HCV infection (8,9). Studies of HCV progression to HCC are expected to provide new insights on the management of this increasing problem and therefore are of great public health interest (10).

The natural progression of HCV infection to hepatitis, cirrhosis and HCC is slow. Chronic hepatitis develops in ~80% of those infected with HCV. Over the course of ≥20 years, 10–30% of HCV carriers develop cirrhosis; patients with cirrhosis have an annual risk of 1–2% for developing HCC (11). The prognosis of patients with HCC remains extremely poor. The currently available systemic therapies demonstrate poor to modest response rates and have not been shown to improve survival in patients with HCC (12). Complete surgical resection and liver transplant are at present the only curative treatment options (13). However, the majority of patients present with advanced unresectable disease not amenable to definitive local therapies (14,15). The slow development and late detection of HCC suggest that the identification of biomarkers of disease progression and early detection represents attractive strategies for potential improvement of the outcome of HCC patients.

Current diagnosis of HCC relies on clinical information, liver imaging and measurement of serum alpha-fetoprotein (AFP). The reported sensitivity (41–65%) and specificity (80–94%) of AFP are not sufficient for early diagnosis, and so additional markers are needed (16,17). The development of effective markers for the diagnosis of HCC could have an impact on HCC-related cancer mortality and significant public health implications worldwide. This is an active area of research with several groups reporting new marker candidates within the last few years (18–21).

The characterization of peptides in serum is a promising strategy for biomarker discovery (22–24). We developed a method for identification of peptides in the enriched low-molecular weight (LMW) fraction of serum based on matrix-assisted laser desorption–ionization time-of-flight (MALDI–TOF) mass spectrometry (MS) (25). In this study, we describe the application of MALDI–TOF MS to the detection of HCC in a cohort of cases and controls recruited from the Egyptian population (5). Our study identified a set of six discriminatory peptide peaks in the sera of HCC patients and controls that could potentially be used for diagnosis of HCC (26). In this paper, we present an analysis of the performance of six candidate markers in an Egyptian study population.

Abbreviations: AFP, alpha-fetoprotein; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; LMW, low-molecular weight; MALDI–TOF, matrix-assisted laser desorption–ionization time-of-flight; MS, mass spectrometry; PSO, particle swarm optimization; SVM, support vector machine.

Materials and methods

Materials

Red-top vacutainer blood collection tubes (BD 366430) were obtained from Becton Dickinson (Franklin Lakes, NJ). C8 magnetic beads, α-cyano-4-hydroxycinaminic acid and MALDI 600 μm AnchorChip were purchased from Bruker Daltonics (Billerica, MA). Microcon 50 kDa ultrafiltration membranes were purchased from Millipore (Bedford, MA). Other chemicals and solvents were purchased from Sigma–Aldrich (St Louis, MO); solvents were of high-performance liquid chromatography grade.

Study population and sample collection

HCC cases and controls were enrolled in collaboration with the National Cancer Institute of Cairo University, Egypt, from 2000 to 2002, as described previously (5). Briefly, adults with newly diagnosed HCC aged 17 and older without a previous history of cancer were eligible for the study. Diagnosis of HCC was confirmed by pathology, cytology, imaging (computer tomography (CT) ultrasound) and serum AFP. Controls were recruited from the orthopedic department of Kasr El Aini Faculty of Medicine, Cairo University. All participants signed informed consent, provided a blood sample and answered a questionnaire with demographic information, personal habits, medical history, occupational history and agricultural activities. The study protocol was approved by the Institutional Review Committee and conformed to the ethical guidelines of the 1975 Helsinki Declaration.

© The Author 2007. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

Blood samples were collected by a trained phlebotomist each day ~10 am and processed within a few hours according to a standard protocol. Aliquots of sera were frozen at −80°C immediately after collection until analysis; all mass spectrometric measurements were performed on twice-thawed sera. Each patient's hepatitis B viral (HBV) and C viral infection status was assessed by enzyme immunoassay for anti-HCV, anti-HBC and HBV surface antigen and by polymerase chain reaction for HCV RNA (5,27). Serum samples of 150 patients from the parent study, consisting of 78 cases and 72 controls matched on age and gender, were analyzed by MALDI–TOF MS as described previously (25,26). Characteristics of this population are summarized in Table I which shows, as expected, increased markers of viral infections (HCV RNA, anti-HCV and anti-HBV) in cancer cases (5,27).

MALDI–TOF MS analysis

We utilized an enrichment procedure to analyze native peptides in the LMW fraction of serum (0.9–5 kDa) using MALDI–TOF MS as described previously (25,28). Briefly, serum samples (15 μl) were desalted on C8 magnetic beads and ultrafiltered in 25% acetonitrile on 50 kDa Microcon membranes. Ultrafiltrates were spotted on AnchorChip MALDI target with α-cyano-4-hydroxycinaminic acid matrix (3.3 mg/ml in 50% acetonitrile). Samples were analyzed on an Ultraflex MALDI–TOF/TOF MS (Bruker Daltonics). Ionization was achieved by irradiation with a nitrogen laser (λ = 337 nm) operating at 20 Hz. An average of 50 shots at each of 20 positions was collected for a total of 1000 shots per spot; the positions were taken in an automated spiral pattern radiating out from the center of the AnchorChip spot. AnchorChip plate locations were calibrated prior to each run. Positive ions were accelerated at 19 kV with 80 ns of pulsed ion extraction delay. Each spectrum was recorded in linear positive mode and was externally calibrated using a standard mixture of peptides. Mass spectra were acquired using the Flex Control and Flex Analysis software (Bruker Daltonics) and raw data were exported as text files for further analysis. To sequence peptides, an exploratory scan from 800 to 5000 Da was performed in the reflectron mode to assign a mass window (~0.5% mass width) for fragmentation and peptide sequencing in the 'LIFT' MS/MS mode. Additional peptide sequencing was carried out on a 4800 MALDI–TOF/TOF™ Analyzer (Applied Biosystems, CA). Analysis was carried out in positive ion mode in both reflector and MS/MS acquisitions with laser repetition rate at 200 Hz. In both reflector and MS/MS mode, the instrument default calibration was used. The positive ion reflector spectra were obtained for the peptide mix in 800–4000 mass range. In MS/MS mode, 2 kV collision energy (with Collision induced dissociation gas ON) was used to fragment the peptides. The database searches for peptide identification were performed using MASCOT Distiller 2.1 (Matrix Science, London, UK). The MS/MS spectra obtained from MALDI–TOF/TOF were searched against the SwissProt human database. No enzyme was considered in these searches and both MS and MS/MS tolerance was 0.3 Da.

Data processing and analysis

Raw spectra are available at http://microarray.georgetown.edu/web/files/carcinogenesis.htm. Analyses were carried out with the Flex Analysis, ClinProTools (Bruker Daltonics), MATLAB (MathWorks, Natick, MA) and SAS (SAS, Cary, NC) software packages. Spectra were processed as described previously (26,29). Briefly, the dimension of each spectrum was reduced from ~136 000 m/z values to 23 846 bins (100 p.p.m. step). Baseline-corrected spectra were normalized by dividing each spectrum by its total ion current; spectra were scaled to an intensity of 100 assigned to the highest peak in the training dataset. For determination of peaks, the dataset was randomly divided into 100 training spectra (50 HCC and 50 control) and 50 blinded testing spectra (28 HCC and 22 control). Peaks were identified as a change from positive to negative slope and nearby peaks within 300 p.p.m. mass were coalesced into a single window to account for drift in m/z location. This procedure identified 264 peak-containing windows in the training spectra; the maximum intensity in each window was used as the variable of interest. A particle swarm optimization (PSO)–support vector machine (SVM) algorithm was applied to the training dataset to select six mass windows for classification of HCC cases and controls. The PSO–SVM algorithm combines two machine learning methods, PSO and SVM, as described previously (26,29). PSO starts with N randomly selected particles and searches for the optimal particle iteratively. Each particle is an m-dimensional vector and represents a candidate solution. An SVM classifier is built for each candidate solution (particle) to evaluate its performance through the cross-validation method. The PSO algorithm guides the selection of potential biomarkers (mass windows) that lead to best prediction accuracy in distinguishing between two groups. The algorithm uses the most-fit particles to contribute to the next generation of N candidate particles. Thus, on the average, each successive population of candidate particles fits better than its predecessor. This process continues until the performance of the SVM classifier converges.

The testing spectra were scaled based on the parameters used for scaling the training spectra and peaks that fall within the selected six mass windows were quantified. Sensitivity and specificity of these six marker candidates were evaluated on the testing dataset. Logistic regression models were used to determine association of the marker candidates and covariates including HCV and HBV viral infections (independent variables) with HCC status (dependent variable).

Table I. IDemographic variables and viral infections

	Cases (n = 78)	Controls (n = 72)	P value
Mean age (SD)	54 (9.1)	52 (12.0)	0.3197
Male gender	57 (73%)	50 (69%)	0.6231
Smokers	43 (55%)	39 (54%)	0.8373
HCV RNA+	62 (80%)	16 (22%)	<0.0001
Anti-HCV+	69 (88%)	24 (33%)	<0.0001
Anti-HBV+	60 (77%)	40 (56%)	0.0050
HBsAg+	5 (6%)	1 (1%)	0.2291

Presence of HCV virus (HCV RNA), HCV antibodies (anti-HCV), HBV surface antigen (HBsAg) and HBV antibodies (anti-HBV) was tested as described in Materials and methods; SD = standard deviation.

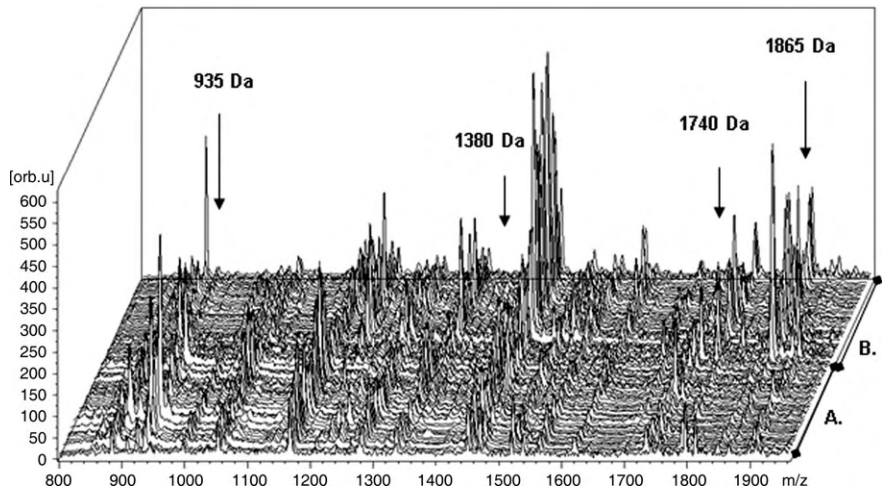


Fig. 1. Overlay of MALDI–TOF spectra in the mass region 0.9–1.9 kDa; Arrows indicate marker candidates selected in this mass region. (**A**) Cases (n = 78); (**B**) controls (n = 72).

Results

Comparison of MALDI–TOF mass spectra of 78 HCC patients and 72 controls matched on age and gender showed marked differences in intensity of several peptide peaks (Figure 1). Serum samples were enriched as described previously (25). Consistency of the differences is visible in the overlay of spectra in Figure 1; differences in the relative ion intensities of average spectra in the mass range of 0.9–5 kDa are presented in supplementary Figure 1 (available at *Carcinogenesis* Online). We selected 6 of 264 peaks for classification of HCC using PSO–SVM computational methods described previously (26,29). Four of the peaks selected for classification of HCC are pointed out in Figure 1 which expands the 0.9–1.9 kDa region; two additional markers were selected in the 2.5 and the 4.1 kDa region, respectively. Intensities of the six marker candidates (MALDI A–F) are summarized in Table II; the values represent baseline-corrected and -normalized intensities (29). The results show that three candidate markers are higher in cases (MALDI B, MALDI C and MALDI D); the remaining three are higher in controls. Overlay of all 150 spectra for two of the peptide peaks in Figure 2 shows the magnitude of the differences. MALDI A is consistently higher in controls ($n = 72$; black traces); MALDI B is higher in cases ($n = 72$; green traces). The smallest difference was observed for MALDI E (Table II). MALDI E was selected based on criteria of prediction accuracy in combination with the other marker candidates even though the difference in mean abundance for some other peptides was greater (see below).

Table III summarizes results of fitting univariate logistic regression models for each marker candidate and important covariate adjusted for matching variables age and gender. The analysis shows that, besides the six marker candidates, several covariates are significantly associated with HCC status. These include HBV and HBC viral infections, residency (urban versus rural) and date of sample collection (see below). In multivariate logistic regression models, we considered each individual marker candidate (MALDI A–F) together with the relevant covariates (HCV RNA, anti-HBC, residency, date of collection, age and gender). As shown in Table IV, the association of marker candidates remains significant after adjustment for all the important covariates. We did not include anti-HCV in the regression models because it is correlated with HCV RNA (correlation coefficient = 0.823). The six selected markers are weakly correlated with each other; the highest correlation coefficient of 0.392 was observed between MALDI C and MALDI D (supplementary Table 1 is available at *Carcinogenesis* Online). This suggests that the markers are independently predictive of HCC and that their combination should have increased prediction accuracy.

The multivariate analysis showed that HCV RNA and date of collection remain significantly associated with the markers. To examine whether MALDI A–F are associated with HCV infection, we fitted univariate logistic regressions with HCV RNA status as the dependent variable and each of the marker candidates in the control population ($n = 72$). The analysis shows that none of the six selected markers is significantly associated with the presence of viral RNA (supplementary Table 2 is available at *Carcinogenesis* Online). We also showed that the HCV presence is not associated with MALDI A–F in cases. This suggests that the observed association of the peptides with HCC is not driven by the viral infection but rather by the presence of the tumor. A potential association of the markers with liver cirrhosis that accompanies most of the HCC cases needs to be addressed in future studies.

Recruitment of cases for the study began prior to the recruitment of controls; controls were subsequently matched to the recruited cases on age and gender. To make sure that the observed association of MALDI A–F is not biased by the time of storage of samples at −80°C, we analyzed the association of MALDI A–F with HCC status in a subset of samples collected simultaneously between January and April of 2002 (supplementary Table 3 is available at *Carcinogenesis* Online). The subset of cases for this analysis is relatively small (22 HCC and 72 controls) but all the marker candidates except for MALDI E remain significantly associated with HCC. This shows that for at least five of the six markers the time of storage does not affect the association with HCC status.

The long-term goal of our study is to identify markers of HCC that would improve the sensitivity and specificity of AFP (16,17). We determined sensitivity and specificity of MALDI A–F on a blinded independent test set of 50 samples (28 HCC cases and 22 controls). The prediction accuracy of individual marker candidates is summarized in Table V. Sensitivity of the markers ranges from 50–96% and

Table II. Descriptive statistics

Marker	m/z window (Da)	Cases ($n = 78$)			Controls ($n = 72$)		
		Mean	Median	SD	Mean	Median	SD
MALDI A	1863–1871	3.7	2.9	3.0	25.9	21.4	21.1
MALDI B	934–938	9.6	4.5	14.5	1.3	0.9	2.5
MALDI C	2529–2536	2.0	1.6	1.3	1.2	1.1	0.5
MALDI D	1737–1744	5.2	3.8	4.5	2.6	2.1	2.0
MALDI E	1379–1381	1.1	1.1	0.5	1.6	1.3	0.9
MALDI F	4086–4098	1.2	0.7	1.2	3.1	3.0	1.9

Normalized peak intensities of marker candidates in cases and controls; SD = standard deviation.

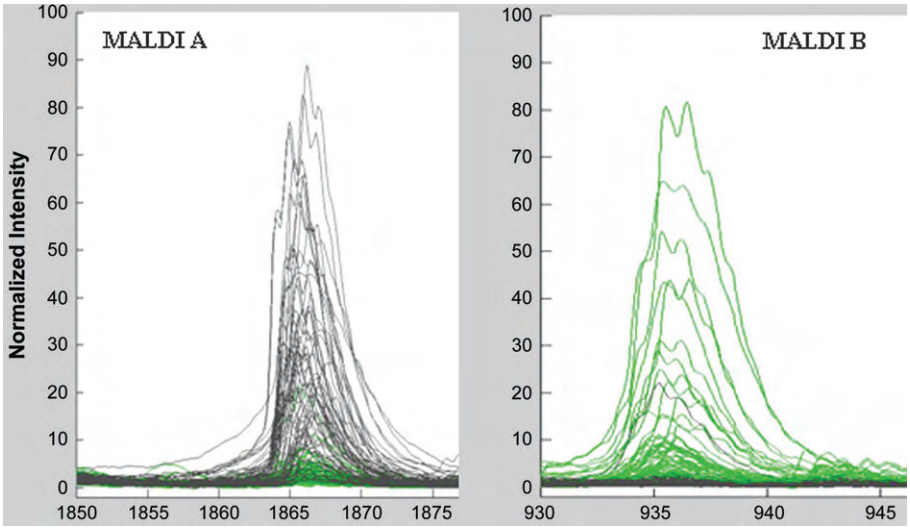


Fig. 2. Overlay of HCC (green traces, $n = 78$) and controls (black traces, $n = 72$) for two selected biomarker candidates; *x*-axis indicates m/z values.

specificity ranges from 36–91%; these values are comparable with the reported performance of AFP (20). Correlation of any of the six peak intensities in 78 HCC cases with AFP was not significant with the correlation coefficients ranging from −0.132 to 0.158; the AFP concentration in controls was not determined. The prediction accuracy in combination with AFP should be evaluated in the future. We were most interested, however, in the performance of a combination of the markers because we expected that a combination of markers would better address the heterogeneity of the disease process. Our analysis shows that the combination of the six selected markers achieved 100% sensitivity and 91% specificity in classification of HCC.

It is important to note that our comparison group represents the general population in Egypt; 33% of the controls carry HCV antibodies and 22% tested positive for HCV RNA. This group of controls is not an arbitrary healthy comparison group and reflects the magnitude of the HCV epidemic in Egypt. Prediction accuracy of the marker candidates is further demonstrated by the receiver operating characteristic curves presented in supplementary Figure 3 (available at *Carcinogenesis* Online). Increased area under the curve for a combined classifier (97% compared with 63–93% for the individual markers as listed in Table V) suggests that a combination of markers is more effective in predicting HCC status than individual markers. Further studies including a cirrhosis comparison group will be needed to evaluate the potential of these diagnostic marker candidates for early detection.

Preliminary sequencing results show that the markers are peptides. Sequencing of MALDI A identified with high probability (MASCOT ion score 99, expectation value 3.5e-6), a fragment of complement C3 with sequence SSKITHRIHWESASLL. Sequencing of MALDI D identified with high probability (MASCOT ions score 131, expect 5.9e-10), a fragment of complement C4a with sequence NGFKSHALQLNNRQI (supplementary Figure 2 is available at *Carcinogenesis* Online). Sequencing of the remaining peptides is under way.

Table III. Association of candidate markers and covariates with HCC

	P value	OR	95% CI	
Anti-HCV	<0.0001	15.3	6.55	35.87
HCV RNA	<0.0001	9.6	4.47	20.70
Anti-HBV	0.0062	2.7	1.32	5.38
HBsAg	0.1535	4.9	0.55	42.66
Residency	0.0017	3.0	1.50	5.82
Smoking	0.8370	1.1	0.56	2.04
Date of collection	<0.0001	10.9	4.60	26.02
MALDI A	<0.0001	0.7	0.65	0.83
MALDI B	<0.0001	2.1	1.52	2.78
MALDI C	<0.0001	3.3	1.87	5.89
MALDI D	<0.0001	1.6	1.26	1.93
MALDI E	<0.0001	0.3	0.16	0.55
MALDI F	<0.0001	0.4	0.32	0.58

Univariate logistic regression adjusted for age and gender ($n = 150$). OR, odds ratio; 95% CI, 95% confidence interval; HBsAg, HBV surface antigen.

Table IV. Association of candidate markers with HCC

	P value	OR	95% CI	
MALDI A	<0.0001	0.70	0.59	0.82
MALDI B	<0.0001	1.76	1.33	2.33
MALDI C	0.0045	2.98	1.40	6.35
MALDI D	0.0011	1.39	1.14	1.69
MALDI E	0.0044	0.33	0.16	0.71
MALDI F	<0.0001	0.45	0.32	0.63

Multivariate logistic regression model controlled for HCV RNA, anti-HBC, residency, date of collection, age and gender ($n = 150$). OR, odds ratio; 95% CI, 95% confidence interval.

Table V. Prediction accuracy of individual markers

	Sensitivity (%)	Specificity (%)	AuROC (%)
MALDI A	96	73	93
MALDI B	82	91	91
MALDI C	50	91	70
MALDI D	64	82	77
MALDI E	71	36	63
MALDI F	82	59	83
All combined	100	91	98

Sensitivity, specificity and AuROC curve derived from an independent test set of 28 cases and 22 controls.

Discussion

AFP is the only marker used currently in the clinic for detection of HCC. Although AFP improves detection of HCC, a significant number of HCC patients present without elevated AFP, and therefore additional markers are needed to increase the sensitivity and specificity of detection (16,30). This study describes an initial validation of six marker candidates selected by MALDI–TOF analysis of enriched LMW fraction of serum.

A recent study identified native peptides associated with cancers of the prostate, breast and bladder by MALDI–TOF/TOF analysis of serum desalted on C8 magnetic beads (22). It was suggested that tumor-associated proteolytic activity is responsible for generation of the diagnostic peptides. Analysis of native peptides begins to provide interesting biological insights and potentially new disease markers (23,31,32). Our study used a newly optimized enrichment procedure for a MALDI–TOF peptidomic analysis of LMW fraction of serum in patients with HCC (25).

We selected 6 of 264 peptide peaks for classification of HCC using previously described computational methods (26). We compared intensities of the six marker candidates in serum samples of 150 patients and found marked differences between HCC cases and controls as demonstrated in Table II and Figures 1–2. The differences are associated with the presence of HCC; the association with HCC remains significant after adjustment for all important covariates (Table IV). We took the following steps to limit the number of false discoveries (33,34). The cases and controls are a representative sample of the Egyptian population (5) with a substantial proportion of controls carrying HCV infection; it is not a convenience sample. A standardized sample collection and processing protocol was used to minimize variability in freeze–thaw cycles and other factors suggested to affect peptide abundance (35). Our analytical methods were optimized to limit variability of the measurements to a mean Coefficient of variation (CV) of ~10% in analysis of 15 replicates of a serum standard (25). We focused our analysis on peaks (as opposed to all observed mass points) and introduced rigorous guidelines for biomarker selection (26,29). The performance of the markers was determined on an independent blinded dataset that was not used for selection of the candidate markers. The identity of two peptides was verified by sequencing.

It is important to note that the selected markers are not the ones with greatest differences between cases and controls based on statistical tests. Each of the six marker candidates is significantly associated with HCC with three of them having higher intensities in cancer patients (Tables II and III) but our computational methods selected markers based on prediction accuracy (29). We expected that a combination of the markers would better classify samples and that interactions may be important in their performance. The results show that the observed correlation coefficients between markers are all smaller than 0.392 which suggests that they are independently predictive of HCC. The individual marker candidates have good

prediction accuracy, in general, comparable with AFP (16). MALDI A has highest sensitivity and MALDI B highest specificity; MALDI E has the lowest prediction accuracy (Table V). We have observed higher sensitivity (100%) and specificity (91%) for the combination of six markers. The prediction accuracy of the six most significantly different peak intensities is lower (96% sensitivity and 82% specificity) which suggests that it is not an optimal criterion for selection of marker combinations.

The presented results provide evidence that our MALDI–TOF analysis of LMW serum fraction identified novel candidate markers of HCC. We expect that these peptides and protein fragments are associated with modified proteolytic activity in the HCC patients; at present, this hypothesis remains a speculation. Our current effort focuses on the sequencing of the candidates and on the analysis of their biological origin. Preliminary sequencing results identified MALDI A and MALDI D as fragments of complement C3 and C4a, respectively (supplementary Figure 2 is available at *Carcinogenesis* Online). Functional implications of these fragments of proteins involved in innate immune response in the serum of patients need further study. Future expansion of the study and comparison with a group of cirrhotic patients is needed to determine if the candidate markers are correlated with cirrhosis and have potential clinical utility for the detection of HCC. The sequenced peptides can be quantified by an appropriate analytical method, for example an isotope dilution mass spectrometric assay.

In summary, we present evidence that six peptide peaks present in LMW fraction of serum are good candidate markers of HCC. The observation should be repeated in an independent set of samples and expanded outside of the Egyptian population. Sequencing of the marker candidates is essential to provide further biological insights and means for accurate quantification.

Supplementary material

Supplementary Tables 1–3 and Figures 1–3 can be found at http://carcin.oxfordjournals.org/

Funding

An Associate Membership from National Cancer Institute's Early Detection Research Network, U.S. Army Medical Research and Material Command, Prostate Cancer Research Program (W81XWH-04-1-0294); National Cancer Institute (R03 CA119288 and R01 CA115625-01A2 to R.G., R03 CA119313 H.W.R. and RO1CA85888 to C.L.).

Acknowledgements

We want to acknowledge Dr Alaa Ismail and Dr Hussain Khaled for invaluable help with design and execution of the studies in Egypt.

Conflict of Interest Statement: None declared.

References

1. Montalto,G. *et al*. (2003) Epidemiology, risk factors, and natural history of hepatocellular carcinoma. *Ann. N. Y. Acad. Sci.*, **963**, 13–20.
2. Parkin,D.M. (2006) The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer*, **118**, 3030–3044.
3. Mizokami,M. *et al*. (2005) Tracing the evolution of hepatitis C virus in the United States, Japan, and Egypt by using the molecular clock. *Clin. Gastroenterol. Hepatol.*, **3**, S82–S85.
4. Deuffic-Burban,S. *et al*. (2006) Expected increase in hepatitis C-related mortality in Egypt due to pre-2000 infections. *J. Hepatol.*, **44**, 455–461.
5. Ezzat,S. *et al*. (2005) Associations of pesticides, HCV, HBV, and hepatocellular carcinoma in Egypt. *Int. J. Hyg. Environ. Health*, **208**, 329–339.
6. Perz,J.F. *et al*. (2006) The coming wave of HCV-related liver disease: dilemmas and challenges. *J. Hepatol.*, **44**, 441–443.
7. Hassan,M.M. *et al*. (2001) The role of hepatitis C in hepatocellular carcinoma: a case control study among Egyptian patients. *J. Clin. Gastroenterol.*, **33**, 123–126.
8. El-Serag,H.B. *et al*. (1999) Rising incidence of hepatocellular carcinoma in the United States. *N. Engl. J. Med.*, **340**, 745–750.
9. Wong,J.B. *et al*. (2000) Estimating future hepatitis C morbidity, mortality, and costs in the United States. *Am. J. Public Health*, **90**, 1562–1569.
10. Di Bisceglie,A.M. *et al*. (2003) Hepatitis C-related hepatocellular carcinoma in the United States: influence of ethnic status. *Am. J. Gastroenterol.*, **98**, 2060–2063.
11. Ikeda,K. *et al*. (1998) Disease progression and hepatocellular carcinogenesis in patients with chronic viral hepatitis: a prospective observation of 2215 patients. *J. Hepatol.*, **28**, 930–938.
12. Llovet,J.M. (2005) Updated treatment approach to hepatocellular carcinoma. *J. Gastroenterol.*, **40**, 225–235.
13. Schwartz,J.M. *et al*. (2003) Treatment of hepatocellular carcinoma. *Curr. Treat. Options Gastroenterol.*, **6**, 465–472.
14. Steel,L.F. *et al*. (2001) A proteomic approach for the discovery of early detection markers of hepatocellular carcinoma. *Dis. Markers*, **17**, 179–189.
15. Lopez,L.J. *et al*. (2004) Hepatocellular carcinoma. *Curr. Opin. Gastroenterol.*, **20**, 248–253.
16. Marrero,J.A. (2005) Screening tests for hepatocellular carcinoma. *Clin. Liver Dis.*, **9**, 235–251.
17. Gupta,S. *et al*. (2003) Test characteristics of alpha-fetoprotein for detecting hepatocellular carcinoma in patients with hepatitis C. A systematic review and critical analysis. *Ann. Intern. Med.*, **139**, 46–50.
18. Filmus,J. *et al*. (2004) Glypican-3 and alphafetoprotein as diagnostic tests for hepatocellular carcinoma. *Mol. Diagn.*, **8**, 207–212.
19. Block,T.M. *et al*. (2005) Use of targeted glycoproteomics to identify serum glycoproteins that correlate with liver cancer in woodchucks and humans. *Proc. Natl Acad. Sci. USA*, **102**, 779–784.
20. Chignard,N. *et al*. (2006) Cleavage of endoplasmic reticulum proteins in hepatocellular carcinoma: detection of generated fragments in patient sera. *Gastroenterology*, **130**, 2010–2022.
21. Lee,I.N. *et al*. (2006) Identification of complement C3a as a candidate biomarker in human chronic hepatitis C and HCV-related hepatocellular carcinoma using a proteomics approach. *Proteomics*, **6**, 2865–2873.
22. Villanueva,J. *et al*. (2006) Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J. Clin. Invest.*, **116**, 271–284.
23. Tammen,H. *et al*. (2005) Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. *Proteomics*, **5**, 3414–3422.
24. Hortin,G.L. *et al*. (2006) Proteomics: a new diagnostic frontier. *Clin.Chem.*, **52**, 1218–1222.
25. Orvisky,E. *et al*. (2006) Enrichment of low molecular weight fraction of serum for mass spectrometric analysis of peptides associated with hepatocellular carcinoma. *Proteomics*, **6**, 2895–2902.
26. Ressom,H. *et al*. (2005) Analysis of MALDI-TOF serum profiles for biomarker selection and sample classification. Proceedings of the IEEE Symposium on Computational Inteligence in Bioinformatics and Computational Biology. La Jolla, CA, 378–384.
27. Abdel-Hamid,M. *et al*. (1997) Optimization, assessment, and proposed use of a direct nested reverse transcription-polymerase chain reaction protocol for the detection of hepatitis C virus. *J. Hum. Virol.*, **1**, 58–65.
28. Tirumalai,R.S. *et al*. (2003) Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics*, **2**, 1096–1103.
29. Ressom,H. *et al*. (2005) Analysis of mass spectral serum profiles for biomarker selection. *Bioinformatics*, **21**, 4039–4045.
30. Zhang,B.H. *et al*. (2004) Randomized controlled trial of screening for hepatocellular carcinoma. *J. Cancer Res. Clin. Oncol.*, **130**, 417–422.
31. Schulz-Knappe,P. *et al*. (2005) The peptidomics concept. *Comb. Chem. High Throughput Screen.*, **8**, 697–704.
32. Hortin,G.L. (2006) The MALDI-TOF mass spectrometric view of the plasma proteome and peptidome. *Clin. Chem.*, **52**, 1223–1237.
33. Ransohoff,D.F. (2005) Bias as a threat to the validity of cancer molecular-marker research. *Nat. Rev. Cancer*, **5**, 142–149.
34. Ransohoff,D.F. (2004) Rules of evidence for cancer molecular-marker discovery and validation. *Nat. Rev. Cancer*, **4**, 309–314.
35. Villanueva,J. *et al*. (2005) Correcting common errors in identifying cancer-specific peptide signatures. *J. Proteome Res.*, **4**, 1060–1072.

Received May 3, 2007; revised June 25, 2007; accepted July 22, 2007

BIOINFORMATICS ORIGINAL PAPER

Vol. 23 no. 5 2007, pages 619–626
doi:10.1093/bioinformatics/btl678

Data and text mining

Peak selection from MALDI-TOF mass spectra using ant colony optimization

H. W. Ressom[1,*], R. S. Varghese[1], S. K. Drake[2], G. L. Hortin[3], M. Abdel-Hamid[4], C. A. Loffredo[1] and R. Goldman[1]

[1]Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC 20057 USA, [2]Critical Care Medicine, NIH, Bethesda, MD 20792 USA, [3]Clinical Chemistry Service, Department of Laboratory Medicine, NIH, Bethesda, MD 20892 USA and [4]Viral Hepatitis Research Laboratory, NHTMRI, Cairo, Egypt

Received on November 19, 2006; revised on December 20, 2006; accepted on January 5, 2007

Advance Access publication January 19, 2007

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Due to the large number of peaks in mass spectra of low-molecular-weight (LMW) enriched sera, a systematic method is needed to select a parsimonious set of peaks to facilitate biomarker identification. We present computational methods for matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) spectral data preprocessing and peak selection. In particular, we propose a novel method that combines ant colony optimization (ACO) with support vector machines (SVM) to select a small set of useful peaks.

Results: The proposed hybrid ACO-SVM algorithm selected a panel of eight peaks out of 228 candidate peaks from MALDI-TOF spectra of LMW enriched sera. An SVM classifier built with these peaks achieved 94% sensitivity and 100% specificity in distinguishing hepatocellular carcinoma from cirrhosis in a blind validation set of 69 samples. Area under the receiver operating characteristic (ROC) curve was 0.996. The classification capability of these peaks is compared with those selected by the SVM-recursive feature elimination method.

Availability: Supplementary material and MATLAB scripts to implement the methods described in this article are available at http://microarray.georgetown.edu/web/files/bioinf.htm

Contact: hwr@georgetown.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

Several proteomic methods show promise in non-invasive screening of accessible fluids such as serum. The characterization of peptides in serum and plasma by mass spectrometry (MS) is one of the promising strategies for biomarker discovery (Tammen *et al*., 2005; Villanueva *et al*., 2006). Biomarker discovery through analysis of mass spectral data requires careful experimental design. It is important to take into account population sampling, matching of controls, protocols for unbiased sample collection, uniform sample preparation methods and appropriate mass spectrometric analysis. Sorace and Zhan (2003) reported the possibility of experimental bias in their assessment of surface enhanced laser desorption/ionization time-of-flight (SELDI-TOF) analysis of ovarian cancer. Ransohoff (2005) indicated that bias will increasingly be recognized as the most important 'threat to validity' that must be addressed in the design, conduct and interpretation of such research. Bias can occur if the cancer and non-cancer groups are handled in systematically different ways, introducing an apparent 'signal' into one group but not the other.

Mass spectra contain true signal and electronic/chemical noise due to contaminants and matrix; this also causes varying baseline (Malyarenko *et al*., 2005). In addition, mass spectra reflect variability in sample preparation and sample degradation. Previous quality-control experiments identified properties of mass spectrometric measurements that must be accounted for in the analysis (Fung and Enderwick, 2002; Yasui *et al*., 2003). The impact of these artifacts can be minimized through data preprocessing steps (smoothing, baseline correction, normalization, peak detection and peak calibration). Previously, we applied these methods to preprocess SELDI-QqTOF and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) spectra (Ressom *et al*., 2005, 2006). We observed that normalization of spectra contributed most to the decrease in observed variability (Orvisky *et al*., 2006). In this study, we add one more data-preprocessing step, which eliminates peaks that are associated with known factors other than disease such as age, gender and smoking status. We call this step *peak screening*. Spectral preprocessing and peak screening is followed by selection of peaks that are associated with the disease under study.

One can distinguish three main approaches for feature selection: *filter, embedded* and *wrapper*. The filter approach is commonly used to select features by applying statistical analyses (e.g. *t*-test, weighting factor, etc.) that recognize differentially expressed peaks between two groups with multiple subjects. The resulting peaks are then used as inputs to a pattern classification algorithm such as support vector machine (SVM). The filter approach provides generic selection of features, not tuned by a given learning algorithm and it is usually fast and easy to interpret. However, it has the following

*To whom correspondence should be addressed.

© The Author 2007. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

limitations: (1) it selects features based on 'relevance' criterion, not on their classification capability; (2) redundant features can exist; (3) features that have strong discriminating power jointly, but are weak individually are ignored.

In the embedded approach, feature selection is part of the training procedure of a classifier. The implementation of this approach depends on the type of the classifier. Guyon *et al.* (2002) proposed an SVM recursive feature elimination (SVM-RFE) algorithm to recursively classify samples with SVM and select features according to their weights in the SVM classifier. SVM-RFE ranks the features once using all samples, and uses the top-ranked features in the succeeding cross-validation for the classifier. This generates a biased estimation of errors and limits the search space by allowing only the top-ranked features as candidates.

In the wrapper method, features are selected by taking into account their contribution to the performance of a given type of classifier. Thus, the wrapper method allows selection of features based on 'usefulness' criterion. It searches for a combination of useful features from the entire set of features. It tends to find features better suited to the classifier, but it also tends to be more computationally expensive than the filter approach. Due to the large number of peaks involved in mass spectra, a systematic method is required to select the best combination of features without examining all possible combinations. Also, wrapper methods are prone to overfitting unless an internal cross-validation method is used during feature selection. Stochastic global optimization methods such as genetic algorithms (GAs), simulated annealing and swarm intelligence (SI) methods are among the ideal candidates for selecting features from a high-dimensional search space. Recently, researchers have investigated the use of these methods to select features including a recent release of ClinProTools that uses GAs to select features for SVM classifiers. In our previous work, we combined particle swarm optimization (PSO) to select SELDI-QqTOF peaks for SVM classifiers (Ressom *et al.*, 2005).

In this article, we present computational methods to identify candidate biomarkers that distinguish hepatocellular carcinoma (HCC) from cirrhosis through MALDI-TOF analysis of enriched low-molecular-weight (LMW) serum samples. As we described previously, denaturing ultrafiltration enriches the LMW fraction of serum by removal of proteins greater than 50 kDa including albumin (Orvisky *et al.*, 2006). The enrichment improves quality of the spectra and allows the analysis of ~300 peptides. Unlike our previous study (Ressom *et al.*, 2006) that identified peaks distinguishing HCC patients from healthy individuals, in this study our goal is to distinguish HCC patients from cirrhotic patients. Cirrhotic patients are at increased risk of developing HCC and monitoring of cirrhotic patients can potentially decrease HCC related mortality rate. To facilitate the identification of the most useful peaks that would lead to discovery of arkers, we applied data preprocessing methods (outlier screening, baseline correction, normalization, peak detection, peak calibration and peak screening). A novel feature selection method that combines ant colony optimization (ACO) with SVM was applied to select a parsimonious set of peaks that achieved high sensitivity and specificity in distinguishing HCC from cirrhosis. The classification capability of the selected peaks is compared with those selected by the SVM-RFE method. Peptide identification and validation of the selected candidate biomarkers is in progress.

2 METHODS

Figure 1 illustrates the steps that we applied to select candidate biomarkers. In the following sections, we briefly describe each of these steps.

2.1 Sample collection

High incidence of viral hepatitis and HCC in Egypt presents a serious health problem. The management of the disease would benefit from identification of biomarkers related to this disease. Serum samples of HCC cases cirrhosis cases, and healthy controls were obtained from Egypt from 2000 to 2002. Controls were recruited among patients from the orthopedic and fracture clinic at the Kasr El-Aini Hospital (Cairo, Egypt) and were frequency-matched to cancer cases by gender, rural versus urban residency and age (Ezzat *et al.*, 2005). All participants signed informed consent, provided a blood sample, and answered a questionnaire with demographic information, personal habits, medical history, occupational history and agricultural activities. Blood samples were collected in red top vacutainer tubes by trained phlebotomist each day around 10 AM and processed within a few hours according to a standard protocol. Aliquots of sera for mass spectrometric analysis were frozen at -80°C immediately after processing. Samples were sub-aliquoted at first thaw and all measurements were performed on samples of second-time thawed serum. Hepatitis C virus (HCV) and hepatitis B virus (HBV) markers in serum samples were determined by EIA for HCV antibodies (anti-HCV), HBV antibodies (anti-HBV) and HBV surface antigen (HBsAg) and by PCR for HCV virus (HCV RNA).

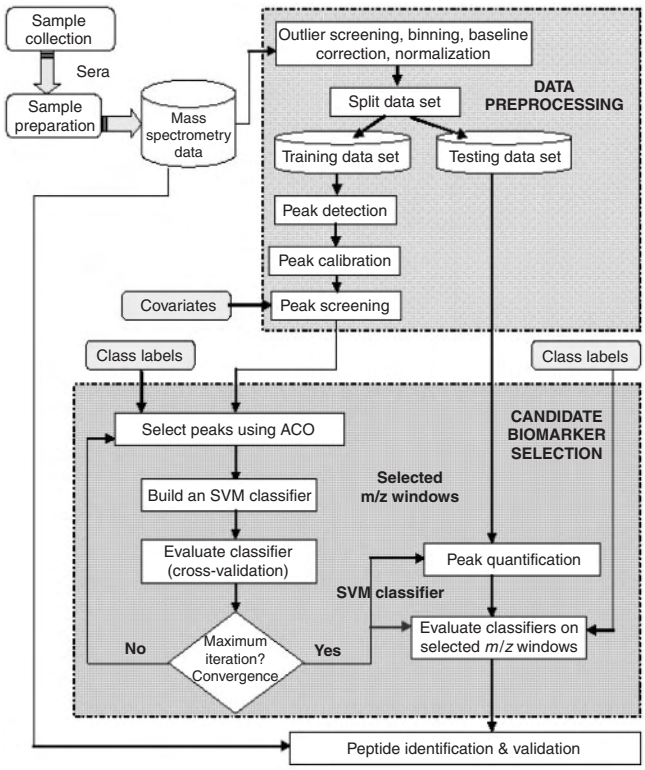


Fig. 1. Methodology for biomarker discovery.

2.2 Sample preparation and data generation

The serum samples were enriched by denaturing ultrafiltration and desalting on C8 magnetic beads (MB) as described earlier (Orvisky *et al*., 2006). The procedure disrupts protein–protein interactions (Tirumalai *et al*., 2003) and allows an efficient recovery of a LMW serum fraction starting with 25 µl of serum. Eluted peptides were mixed with a matrix solution (3 mg/ml α-cyano-4-hydroxycinaminic acid in 50% acetonitrile/0.1% trifluoracetic acid), spotted onto AnchorChip target (Bruker Daltonics, Billerica, MA, USA) and analyzed using an Ultraflex MALDI TOF/TOF mass analyzer (Bruker Daltonics). Each spectrum was detected in linear positive mode and was externally calibrated using a standard mixture of peptides. Using the MALDI-TOF MS, we generated a total of 277 spectra, of which 62 were replicate spectra generated from a serum of one healthy individual. The replicated spectra are useful to assess the run-to-run reproducibility of our sample preparation method and the MALDI-TOF MS technology. The remaining 215 were generated using 84 sera from HCC patients 51 sera from cirrhotic patients, and 80 sera from healthy individuals. Each spectrum consisted of about 136 000 m/z values with the corresponding ion intensities over the mass range 0.9–10 kDa.

2.3 Data preprocessing

Our mass spectral preprocessing method began with outlier screening, where spectra whose total ion current differed by more than two standard deviation (SD) from the median total ion current. This method excluded 14 spectra out of 277 MALDI-TOF spectra generated in this study. Our subsequent analyses were done using the remaining 263 spectra (62 replicate spectra from a serum of one individual, 78 from patients with HCC, 51 from cirrhotic patients and 72 from healthy individuals).

The spectra were binned to reduce their dimension. We used a bin size of 100 ppm. The mean of the intensities within each bin was used as the protein expression variable (Villanueva *et al*., 2004). This method reduced the dimension of each spectrum from about 136 000 m/z values (over the mass range 0.9–10 kDa) to 23 846 m/z bins.

We used the 62-binned replicate spectra to assess the run-to-run reproducibility of our experiment. We transformed each intensity value in the 23 846 m/z bins by computing the base-two logarithm and found the mean log intensity value and SD of the 62 replicate spectra. The Coefficient of Variation (CV) of the 62 log-transformed intensity values ranged between 4.1 and 22.9% with a mean value of 10.5%.

For each of the other 201-binned spectra (78 HCC, 51 cirrhosis and 72 normal), we estimated the baseline by obtaining the minimum value within a shifting window size of 50 bins and a step size of 50 bins. Spline approximation was applied to regress the varying baseline. The regressed baseline was smoothed using the lowess smoothing method. The resulting baseline was subtracted from the spectrum. Then, each spectrum was normalized by dividing by its total ion current.

Prior to peak detection, we split the 78 HCC and 51 cirrhosis spectra into two data sets, training and testing. We used the training data set that consists of 30 HCC and 30 cirrhosis spectra for peak detection and calibration. The 72 spectra from healthy individuals were used for peak screening. The testing data set that consists of 48 HCC and 21 cirrhosis spectra was set aside for later evaluation as a blind data set. A peak is defined as the location of a mass point, where the sign of the slope of the intensity level (ion count) changes from positive to negative and a reasonable intensity level is measured at the mass point. To address the latter requirement, we defined a threshold line and discarded all peaks below it. To accommodate the fact that the noise level decreases as the mass increases, we defined a threshold line that is higher in the low-mass region than in the high-mass region. This is accomplished by first scaling all spectra to an overall maximum intensity of 100 and then discarding peaks under a threshold line that linearly decreases from intensity level 1–0.5 in the mass range 0.9–10 kDa.

Peak calibration was done based on the method proposed by Coombes *et al*. (2004). All the detected peaks were aligned by coalescing neighboring peaks within and across all training spectra into m/z windows. First, we selected peaks above a threshold line that decreases linearly from 4 to 1%. Then, we combined these peaks if they differed in location by at most 7 bins or at most 0.09% relative mass. Following this, we considered peaks with intensities between the threshold line that decreases from 4 to 1% and another threshold line, which decreases from 1 to 0.5%. These peaks were added into previously identified m/z windows if they fell within 7 bins or at most 0.09% relative mass. Note that this step may increase the width of an m/z window if a peak is added from outside, otherwise the m/z window size remains unchanged except that the number of peaks in that window will increase. We retained m/z windows that consisted of peaks from at least five spectra and discarded the rest.

Finally, we examined each peak for a potential association with covariates such as age, gender, smoking status, viral infection and residency. This analysis was performed on the samples from healthy individuals to unambiguously identify peaks associated to the covariates. Two approaches were considered in this analysis. The first approach fitted a logistic regression model where the independent variables were the intensities of a given peak across all normal samples with the status of a given covariate as the dependent variable. All covariates in this study have binary values including age (young versus. old). The association of every peak to each covariate was determined on the basis of the corresponding statistical significance ($P<0.05$) in fitting a logistic regression model. In the second approach, we fitted a linear regression model for each peak. The peak intensity was the dependent variable of the linear regression model, while all the covariates were used as independent variables. For each peak, covariates that were found to be predictors of the peak intensity with a statistical significance of $P<0.05$ were identified. A peak was removed from subsequent analyses, if the above two approaches showed statistically significant association with at least one of the covariates.

2.4 Peak selection

Support Vector Machines are learning kernel-based systems that use a hypothesis space of linear functions in high-dimensional feature spaces. In classification problems that involve two classes, linear SVMs search for the optimal hyperplane that maximizes the margin of separation between the hyperplane and the closest data points on both sides of the hyperplane. Thus, parameters of SVMs are determined on the basis of structural risk minimization, not error-risk minimization. Thus, they have the tendency to overcome the overfitting problem. In high-dimensional data classification problems, SVMs have proven themselves as one of the pattern classification algorithms with great generalization ability.

Ant colony optimization studies artificial systems that take inspiration from the behavior of real ant colonies (Dorigo *et al*., 1999). The basic idea of ACO is that a large number of simple artificial agents are able to build good solutions to solve hard combinatorial optimization problems via low-level based communications. Real ants cooperate in their search for food by depositing chemical traces (pheromones) on the ground. Artificial ants cooperate by using a common memory that corresponds to the pheromone deposited by real ants. The artificial pheromone is accumulated at runtime through a learning mechanism. Artificial ants are implemented as parallel processes whose role is to build problem solutions using a constructive procedure driven by a combination of artificial pheromone and a heuristic function to evaluate successive constructive steps. Our motivation for using ACO for feature selection is due to its efficiency and capability in identifying a set of interacting variables that are useful for classification.

Also, ACO allows the integration of prior information into the algorithm for improved feature selection.

Through the probability function given below, each ant picks n sets of distinct features from L candidate peaks:

$$P_i(t) = \frac{(\tau_i(t))^{\alpha}\eta_i^{\beta}}{\sum_i (\tau_i(t))^{\alpha}\eta_i^{\beta}} \quad (1)$$

where $\tau_i(t)$ is the amount of pheromone trail at time t for the feature represented by index i; η_i represents prior information (e.g. univariate t-statistic) for the feature represented by index i; α and β are parameters that determine the relative influence of pheromone trail and prior information.

At $t=0$, $\tau_i(t)$ is set to a constant for all features. Thus, at the first iteration, each ant chooses n distinct features (a trail) from L features with probabilities proportional to the existing prior knowledge. Let S_j be the jth ant consisting of n distinct features. Depending on the performance of S_j, the amount of pheromone trail for S_j is updated. The performance function is evaluated on the basis of disease state classification capability of each S_j. We use the features in S_j to build a classifier and estimate the classification accuracy through the cross-validation method. The amount of pheromone trail for each feature in S_j is updated in proportion to the corresponding classification accuracy using

$$\tau_i(t+1) = \rho \cdot \tau_i(t) + \Delta\tau_i(t) \quad (2)$$

where ρ is a constant between 0 and 1, representing the evaporation of pheromone trails; $\Delta\tau_i(t)$ is an amount proportional to the classification accuracy of S_j, $\Delta\tau_i(t)$ is set to zero if $f_i \notin S_j$. This update is made for all N ants ($S_1, \ldots, S_N$). Note that at $t=0$, $\Delta\tau_i(t)$ is set zero for all features. The updating rule allows trails that yield good classification accuracy to have their amount of pheromone trail increased, while others gradually evaporate. As the algorithm progresses, features with larger amounts of pheromone trails and strong prior information influence the probability function to lead the ants towards them.

ACO-SVM combines ACO and SVM to select peaks that are useful for SVM classification of two groups. ACO starts with a population of N peak sets, where each peak set consists of a pre-specified number (n) of distinct peaks. Each peak is selected from a given set of candidate peaks (L) based on its probability function described previously in Equation (1). SVM classifiers are then built for each peak set and the performance of the peak set in distinguishing the two groups is evaluated through the 4-fold cross-validation method. Using Equation (2), we update the amount of pheromone trail for each peak in proportion to the classification accuracy of the peak set, in which the peak is involved. The goal is to provide those peaks that can lead to improved classification accuracy with better probability of being selected in subsequent iterations. In the following, we illustrate the hybrid ACO-SVM algorithm through an example, where we applied the algorithm to select five peaks from a total of 228 candidate peaks.

In this example, the parameters of ACO-SVM were set to the following values: $N=50$, $L=228$, $n=5$, $\alpha=1$, $\beta=1$ and $\rho=0.1$. Also, the weighting factor proposed by Golub *et al.* (1999) was considered as prior information. Hence, $\eta_i = |\mu_{1i} - \mu_{2i}|/(\sigma_{1i} + \sigma_{2i})$ was used in Equation (1), where μ_{1i} and μ_{2i} represent the mean intensity of peak i in the first and the second group, respectively. Similarly σ_{1i} and σ_{2i} denote the corresponding SDs. We ranked the peaks from 1 to 228 based on decreasing order of weighting factor. Figure 2 depicts the location of each of these peaks in a two-dimensional space. Note that the dimension of the search space and the order of the peaks in the search space do not play a role, because the objective here is to maximize classification accuracy, not the distance between points. At the 1st iteration, each ant chose five peaks from 228 peaks based on Equation (1) that assigns each peak a probability of being selected. This probability function takes only the weighting factor into account during

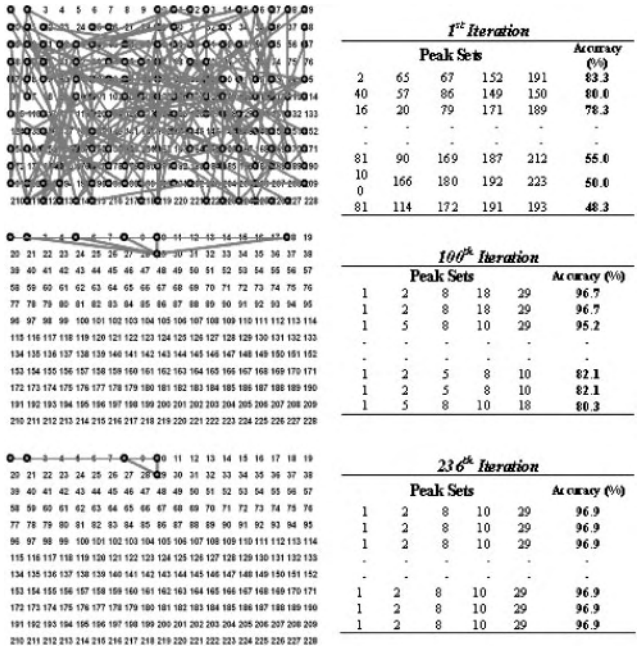


Fig. 2. Pheromone trails for 50 ants at the 1st iteration (top), 100th iteration (middle) and 236th iteration (bottom).

the first iteration. This is because each peak has the same amount of pheromone trail initially. In subsequent iterations, the product of amount of pheromone trail and weighting factor constitute the probability function. Figure 2 (top left figure) shows all 50 ants chosen at the first iteration, where each ant is represented by a trail of five connected circles. The classification capability of each peak set was estimated by the 4-fold cross-validation method. The top right table in Figure 2 shows the ants (peak sets) for the 1st iteration and their corresponding classification accuracy (only the top and bottom three ants are shown, sorted in decreasing order of classification accuracy). $\Delta\tau_i(t)$ in Equation (2) is determined for each peak in proportion to the classification accuracy of the peak set it belongs to. Based on this value, the amount of pheromone trail for each selected peak will be increased (Equation 2). This implies that the probability function for these peaks will increase proportionally, thereby increasing the chance of being selected in the next iteration. Peaks that were not selected will have zero $\Delta\tau_i(t)$. Thus, their corresponding amount of pheromone trail will decrease as a result of the evaporation constant (ρ). The algorithm searches iteratively for more useful peaks guided by the estimated classification capabilities of the peak sets. At the 100th iteration, ants seemed to converge to some trails (Fig. 2, center left figure), primarily to trails that consist of features ranked favorably by weighting factor. At the 236th iteration (bottom left figure), all ants (50 peak sets) converged to one trail that goes through the peaks labeled by indices 1, 2, 8, 10 and 29. Note that at each evaluation, the training spectra are reshuffled. Thus, a repeat of the cross-validation test for the same peak set may result in a different estimate of classification accuracy. In this example, the classification accuracy range across the 50 ants improved from 48.3–83.3% at the 1st iteration to 80.3–96.7% at the 100th iteration, and converged to 96.9% accuracy at the 236th iteration. The algorithm was run for 500 iterations and no change in the peak set was observed after the 236th iteration. The algorithm took about 6.6 min on a Linux machine with dual processor of each Intel Xeon 3 GHz and 6 GB RAM to complete 236 iterations.

3 RESULTS

3.1 Data preprocessing

By applying our peak detection and calibration methods to the training data set that consists of 30 HCC and 30 cirrhosis spectra (binned, baseline corrected and normalized), we obtained 249 m/z windows from 23 846 bins. For each spectrum, the maximum intensity within each window was found, yielding a 249×60 data matrix. We quantified the peaks in the 72 spectra from healthy individuals at the 249 m/z windows, yielding a 249×72 data matrix. Of the 249 peaks, 21 were found to be associated to at least one of the covariates ($P < 0.05$) in both peak screening methods described previously, linear and logistic regression. Table 1 presents the demographic and viral infection data for the samples from healthy individuals and the number of peaks associated with each covariate. Some of the peaks were associated to more than one covariate.

3.2 Peak selection

The aim is to identify candidate biomarkers that distinguish HCC samples from cirrhosis samples. We used the ACO-SVM algorithm to select peaks (m/z windows) from the training HCC and cirrhosis spectra. Only peaks that are not associated with known covariates were considered. Thus, after removing 21 peaks, a training matrix of 228×60 was considered for peak selection. We ran the algorithm 100 times, where each run selected five m/z windows out of 228 m/z windows. A 4-fold cross-validation was used to estimate the classification accuracy. At each evaluation, the training spectra were reshuffled. Thus, a repeat of the cross-validation for the same peak may result in a different estimate of classification accuracy. Each run consisted of a maximum of 500 iterations. Figure 3A and B depicts the frequency of occurrence of the m/z windows selected in 100 runs, sorted by frequency and by weighting factor, respectively. Eight m/z windows were selected in more than 20% of the runs.

To evaluate the generalization capability of the peaks and the SVM classifier determined by the training data set, we used them to classify the spectra in the blind validation set, i.e. the testing spectra that were set aside during the process of data preprocessing, peak selection and building the SVM classifier. We binned, baseline corrected and normalized the testing spectra in the same way as the training spectra. Note that the testing spectra were scaled based on the parameters that were used to scale the training spectra. Figure 4 depicts the receiver operating characteristic (ROC) curves and their corresponding area under the ROC (AUROC) for the eight m/z windows, which were evaluated separately and combined in distinguishing HCC patients from cirrhotic samples in the testing data set. This figure demonstrates the generalization capability of the selected peaks and the SVM classifier in a blind validation set and the advantage of a panel of biomarkers in improving the AUROC.

To further evaluate the usefulness of the small set of m/z windows selected by ACO-SVM, we built three SVM classifiers using three sets of features (the 23 846 m/z bins, 228 m/z windows, and the selected eight m/z windows). Note that each classifier was built using the training spectra and evaluated on the testing spectra in distinguishing HCC from cirrhosis. Figure 5 shows that the AUROC for the SVM classifier with only eight m/z windows performed as good as those that used all m/z bins or all m/z windows. Figure 6 depicts the box plots for the eight m/z windows using peaks from both training and testing spectra (i.e. 78 HCC and 51 cirrhosis spectra).

Table 1. Demographic variables and viral infections and the number of peaks associated for the 72 samples from healthy individuals

Covariate	Number of samples (percentile)	Number of peaks associated
Age < 52	40 (56%)	4
Male: gender	50 (69%)	0
Urban	39 (54%)	1
Smokers	39 (54%)	1
HCV RNA+	16 (22%)	4
anti HCV+	24 (33%)	1
anti HBV+	40 (56%)	11

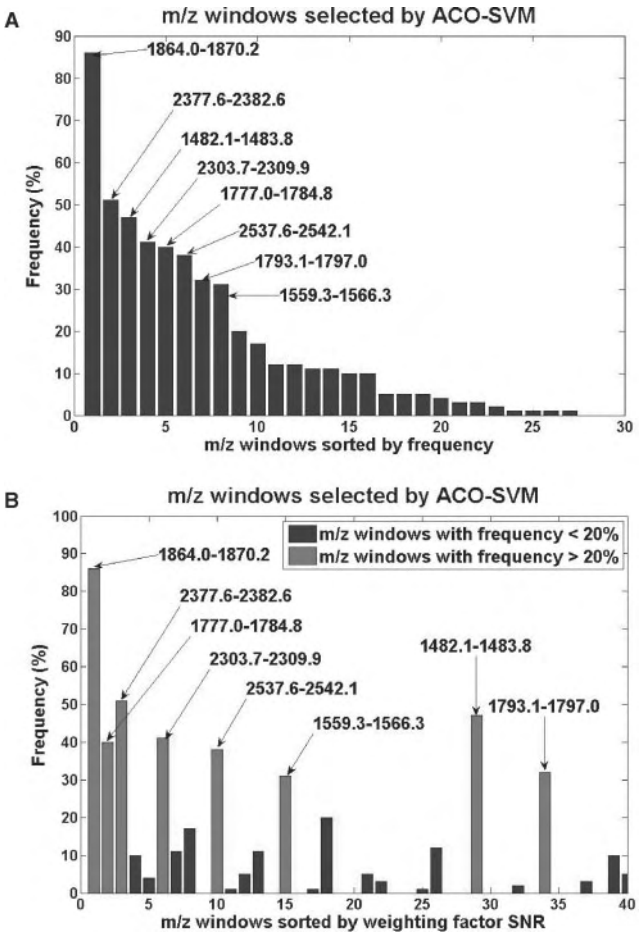


Fig. 3. Frequency of occurrence of peaks selected by ACO-SVM in 100 runs. Panel **A**: peaks sorted by frequency. Panel **B**: peaks sorted by weighting factor.

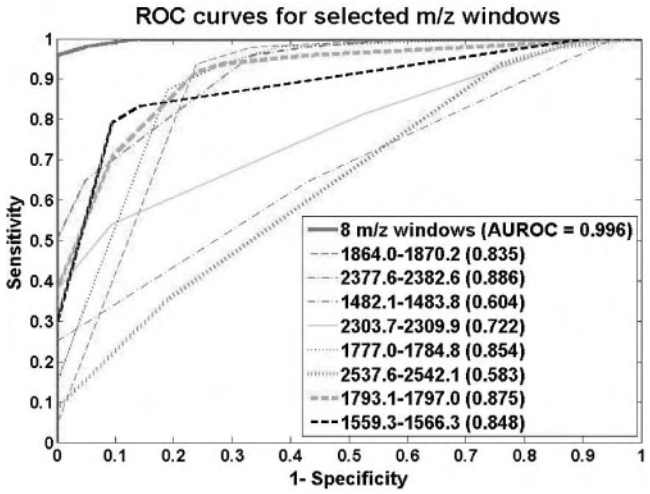


Fig. 4. ROC curves of each of the eight m/z window separately and combined based on the testing spectra (i.e. a blind validation set of 38 HCC and 21 cirrhosis samples that was not involved in peak detection, peak calibration, peak selection or building the SVM classifiers). A colour version of this figure is available as supplementary material.

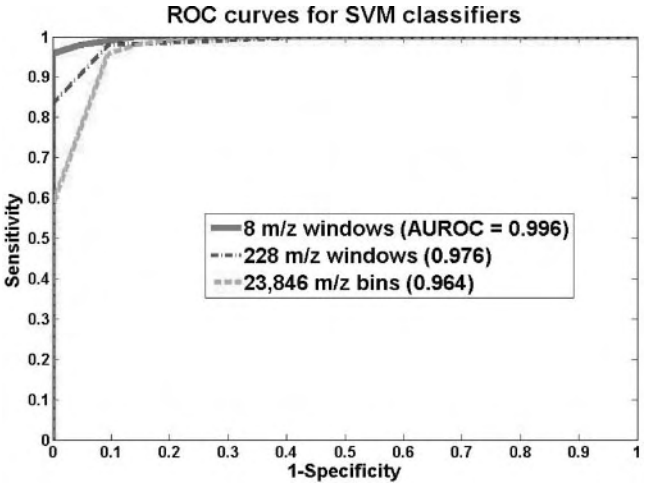


Fig. 5. ROC curves of three SVM classifiers (all bins, all m/z windows and four m/z windows) based on the testing spectra (i.e. a blind validation set of 38 HCC and 21 cirrhosis samples that was not involved in peak detection, peak calibration, peak selection or building the SVM classifiers). A colour version of this figure is available as supplementary material.

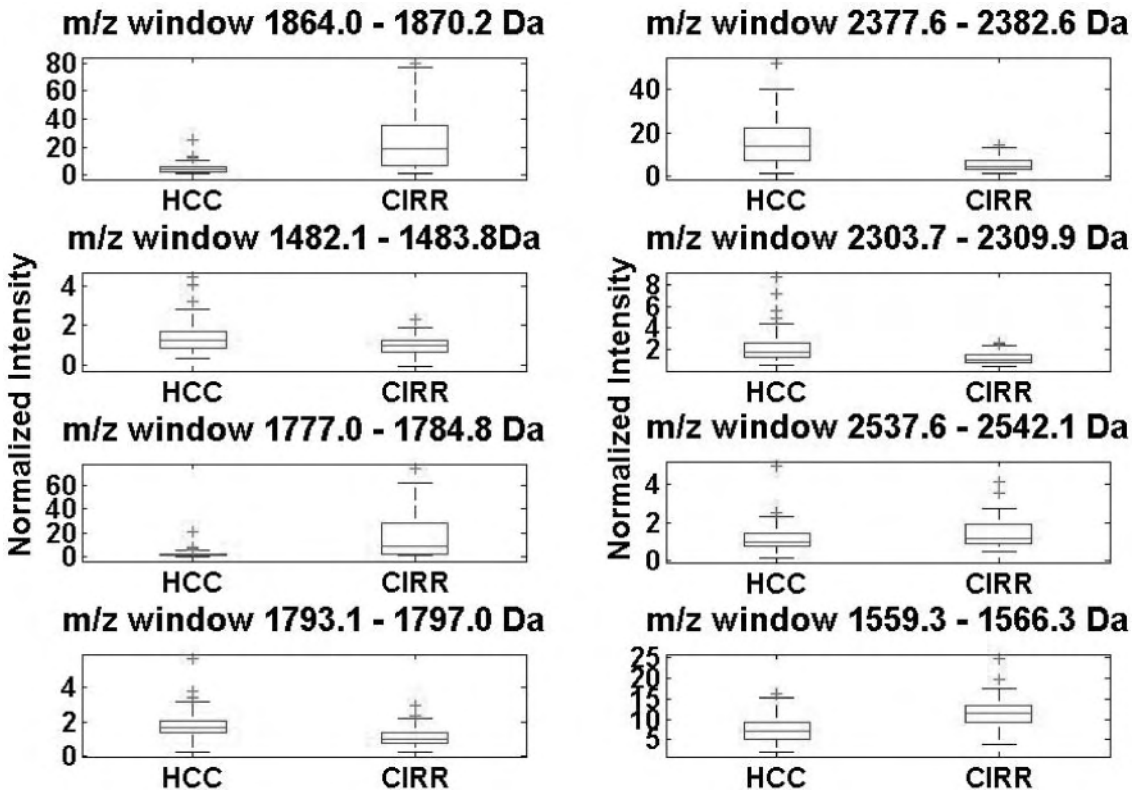


Fig. 6. Box plots for eight m/z windows selected by ACO-SVM (78 HCC and 51 cirrhosis samples).

For comparison, we used the SVM-RFE method to select eight m/z windows from the 228 candidate peaks. Table 2 presents the top eight m/z windows selected by SVM-RFE, weighting factor and ACO-SVM. Four overlaps exist between weighting factor and ACO-SVM, of which two were also selected by the SVM-RFE. These two m/z windows are the top two m/z windows selected by ACO-SVM. The table also shows the sensitivity and specificity achieved by these

Table 2. Eight m/z windows in the order of their weighting factor, frequency of occurrence in ACO-SVM runs and SVM-RFE ranking. The sensitivity and specificity are calculated based on the testing spectra

Weighting factor	ACO-SVM	SVM-RFE
Sensitivity: 88%	Sensitivity: 94%	Sensitivity: 85%
Specificity: 76%	Specificity: 100%	Specificity: 95%
1864.0–1870.2	1864.0–1870.2	3768.9–3777.2
1777.0–1784.8	2377.6–2382.6	1018.3–1022.1
2377.6–2382.6	1482.1–1483.8	1864.0–1870.2
1689.6–1695.9	2303.7–2309.9	2377.6–2382.6
1349.2–1355.9	1777.0–1784.8	933.2–947.1
2303.7–2309.9	2537.6–2542.1	2550.3–2556.4
1313.3–1317.1	1793.1–1797.0	1277.0–1282.5
6593.3–6690.9	1559.3–1566.3	2753.6–2759.9

windows in distinguishing HCC from cirrhosis in the testing data set.

Finally, we examined the consistency of the ACO-SVM algorithm in multiple runs and the impact of its free parameters such as the number of variables selected in each run (n) and the number of ants (N) by running the algorithm for $n = 3,5,7$ and 9 with $N = 25$, 50 and 100. Each combination was run 25 times. Figure 7A and B shows the frequency plot for the total 300 runs, where nine m/z windows were selected in more than 20% of the runs. Seven of the eight m/z windows found earlier were selected again along with two new m/z windows. The sensitivity and specificity of an SVM trained with the nine m/z windows calculated on the testing spectra were both 90%.

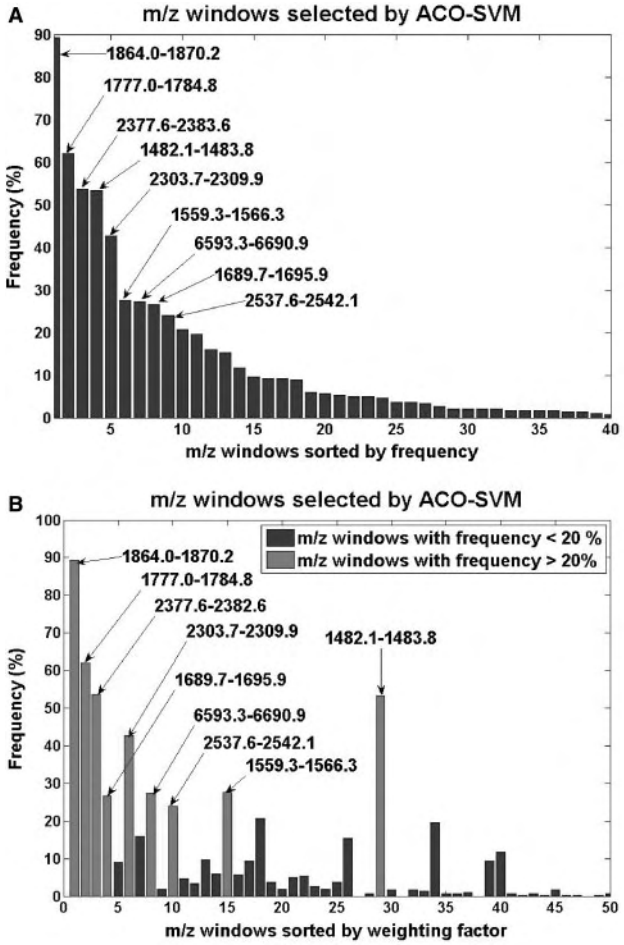


Fig. 7. Frequency of occurrence of peaks selected by ACO-SVM in 300 runs. Panel **A**: peaks sorted by frequency. Panel **B**: peaks sorted by weighting factor.

4 CONCLUSIONS

We present a novel algorithm that combines ant colony optimization with support vector machines to select the most useful peaks from a large number of candidate peaks. The candidate peaks were derived by preprocessing MALDI-TOF spectra of low-molecular-weight enriched sera. Prior to peaks selection, we removed peaks associated to covariates such as age, gender, residency, smoking and viral infection. A small set of peaks selected by the hybrid ACO-SVM algorithm achieved high sensitivity and specificity in distinguishing cirrhotic patients from patients with HCC. Identification of the peptides that the selected peaks represent is in progress. Following peptide identification, the authors plan to perform independent laboratory experiments to validate the candidate biomarkers.

ACKNOWLEDGEMENTS

This work was supported in part by the NCI grant 2R01CA85888 awarded to C.A.L. and by the US Army Medical Research and Material Command's Prostate Cancer Research Program grant W81XWH-04-1-0294, NCI grant 1R03CA119288-01A1 and Associate Membership from NCI's Early Detection Research Network awarded to R.G.

Conflict of Interest: none declared.

REFERENCES

Coombes,K.R., Tsavachidis,S., Morris,J.S., Baggerly,K.A., Hung,M.C., Kuerer, H.M. Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. Technical Report UTMDABTR-001-04, The University of Texas M.D. Anderson Cancer Center, 2004, Available at http://www.mdanderson.org/pdf/biostats_utmdabtr-001-04.pdf

Dorigo,M. *et al.* (1999) Ant algorithms for discrete optimization. *Artif. Life*, **5**, 137–172.

Ezzat,S. *et al.* (2005) Associations of pesticides, HCV, HBV, and hepatocellular carcinoma in Egypt. *Int. J. Hyg. Environ. Health*, **208**, 329–339.

Fung,E.T. and Enderwick,C. (2002) ProteinChip clinical proteomics: computational challenges and solutions. *Biotechniques*, 32 (Suppl), 34–41.

Golub,T.R. *et al.* (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, **286**, 531–537.

Guyon,I. *et al.* (2002) Gene selection for cancer classification using support vector machines. *Machine Learning*, **46**, 389–422.

Malyarenko,D.I. *et al.* (2005) Enhancement of sensitivity and resolution of surface-enhanced laser desorption/ionization time-of-flight mass spectrometric records for serum peptides using time-series analysis techniques. *Clin. Chem.*, **51**, 65–74.

Orvisky,E. *et al*. (2006) Enrichment of low molecular weight fraction of serum for MS analysis of peptides associated with hepatocellular carcinoma. *Proteomics*, **6**, 2895–2902.

Ransohoff,D.F. (2005) Bias as a threat to the validity of cancer molecular-marker research. *Nat. Rev. Cancer*, **5**, 142–149.

Ressom,H.W. *et al*. (2005) Analysis of mass spectral serum profiles for biomarker selection. *Bioinformatics*, **21**, 4039–4045.

Ressom, H.W. *et al*. (2006) Biomarker identification and rule extraction from mass spectral serum profiles. In *Proceedings of the IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, Toronto, ON, pp. 164–170.

Sorace,J.M. and Zhan,M. (2003) A data review and re-assessment of ovarian cancer serum proteomic profiling. *BMC Bioinformatics*, **4**, 24.

Tammen,H. *et al*. (2005) Peptidomic analysis of human blood specimens: comparison between plasma specimens and serum by differential peptide display. *Proteomics*, **5**, 3414–3422.

Tirumalai,R.S. *et al*. (2003) Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics*, **2**, 1096–1103.

Villanueva,J. *et al*. (2004) Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. *Anal. Chem*., **76**, 1560–1570.

Villanueva,J. *et al*. (2006) Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J. Clin. Invest*., **116**, 271–284.

Yasui,Y. *et al*. (2003) A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics*, **4**, 449–463.